

Project Deliverable Report

D4.1 – Positioning Design

Work Package	4		
Task	1, 2		
Date of delivery	Contractual: 01-10-2009	Actual: 05-10-2009	
Code name	D4.1	Version: 2	Draft <input type="checkbox"/> Final <input checked="" type="checkbox"/>
Type of deliverable	Report		
Security (distribution level)	Public		
Contributors	Gaston Burek, Dale Gerdemann, Adriana Berlanga, Els Boshuizen, Isobel Braidman, Alisdair Smithies, Fridolin Wild, Petya Osenova, Kiril Simov, Gillian Armit, Stefan Trausan-Matu		
Authors (Partner)	UTU, OUNL, UNIMAN, IPP-BAS, PUB-NCIT		
Contact Person	Dale Gerdemann		
WP/Task responsible	UTU		
EC Project Officer	Ms. M. Csap		
Abstract (for dissemination)	<p>This report explains the issues of positioning learners in a knowledge domain, to recommend learning materials to follow, and diagnosing learners' conceptual development, to provide formative feedback and recommend remedial actions. It also discusses how Language Technologies can be used to perform these two tasks in a (semi) automatic way. To this end, the report presents an outline of the existing language technologies, and resources for the analysis of learner texts (task 4.1), and an overview of the process of acquiring and analysing data for measuring conceptual development (task 4.2).</p> <p>Furthermore, it presents a description of an initial set of experiments and test results, a plan for extending existing tools for the positioning task, as well as a description service that will be validated in the next cycle of the project.</p>		
Keywords List	Learner Positioning, Diagnosing Conceptual Development, ePortfolio, Latent Semantic Analysis, Knowledge Rich Approaches, Language Technologies, LTfLL		

Table of Contents

Executive summary	3
Work package 4.1: Positioning	3
Work package 4.2: Conceptual development	4
1. Introduction.....	6
1.1 Purpose of document	6
1.2 Project goals	6
1.3 Tasks of work package 4 and their relation to other work packages	6
1.4 Positioning and language technologies (WP4.1)	7
1.5 Conceptual development of the learner (WP4.2).....	8
1.6 Relationship between WP4.1 and WP4.2	9
1.7 Overview of this report.....	9
2. Theoretical background	11
2.1 Introduction	11
2.2 Theoretical basis of learning	11
2.3 Knowledge creation theories	12
2.4 Theories which model the gaining of expertise and their relevance to professional development.....	16
2.5 Knowledge restructuring.....	17
2.6 Professional learning	18
2.7 Learning processes	19
3. Educational domains included in WP4	20
3.1 Medicine as a model.....	20
3.2 Pedagogic domain: Reflective and critical learning in medicine – role of portfolio ...	20
3.3 The role of ePortfolios in Computer Science education.....	22
4. Texts available for analysis	23
4.1 Introduction.....	23
4.2 Text material generated in Manchester University medical curriculum	23
5. Positioning the learner (Task 4.1)	25
5.1 Introduction.....	25
5.2 Knowledge poor based positioning	25
5.3 Knowledge rich approaches for positioning the learner	33
6. Diagnosing conceptual development (task 4.2).....	42
6.1 Outputs of WP4.2.....	42
6.2 Research problems.....	42
6.3 Introduction to studies	43
6.4 Comparison of existing concept mapping tools - which tools to use?	44
6.5 Studies on potential reference models using Leximancer and Pathfinder.....	49
6.6 Conclusions and next steps	51
6.7 Design of the WP4.2 service.....	52
7. Conclusions.....	54
7.1 WP4.1	54
7.2 WP4.2	54
References	56

Executive summary

LTfLL work package 4 focuses on two independent but connected issues:

- Determining the learner's position with regard to learning materials to provide the learner with the 'best' suitable material to achieve their learning goals (WP4.1).
- Determining the conceptual development of a learner related to a particular expertise area, to provide them with formative feedback (WP4.2).

Work package 4.1: Positioning

The WP4.1 scenario is associated with the 'building collaborative knowing' part of the Stahl cycle that integrates the LTfLL project. Brown and Duguid (2001) argue that communities of practice develop knowledge and share that knowledge within the community's participants according to local communication patterns. To develop the positioning system, we will identify natural language expressions characterising the use of language within specific communities of practice, using knowledge poor and knowledge rich approaches e.g. LSA, ontology supported sentiment analysis. Our work in the first phase of the project is described below.

Knowledge poor approach:

Method: The knowledge poor approach restricts itself to analysing only learners' texts and modelling only experts' texts. We use techniques of text categorization e.g. LSA with a traditional bag-of-words model and a novel bag-of-phrases model, where the phrases are extracted using suffix arrays as in Yamamoto and Church (2001). Phrases are weighted according to their probability of occurring predominately in high quality expert texts (representative of an expert community of practice) or low quality non-expert texts.

Results: For comparing and finding prototypical expert texts, phrase-based LSA results generally provide improvement over the traditional bags-of-words LSA results. The phrase weighting approach has been successfully used to extract synonymous pairs (e.g. "drug charts" vs "prescription charts") differing only in standards of usage in communities of practice.

Conclusions and future work: The results so far have been positive for the data we have used. Therefore, we would like to generalise by using texts generated within different communities of practice. As the new texts will present different linguistic features, we plan to test alternative configurations (higher degree of distinctiveness, etc) for the phrase weighting and extraction. Using larger training data sets will allow us to afford the analysis of more distinctive phrases without facing unmanageable levels of sparseness. More distinctive phrases will allow an improvement in the characterisation of language usage, which will help both with the qualitative and quantitative feedback to the user.

Knowledge-rich approach:

According to Wenger (2001), communities of practice produce and share knowledge artifacts that need effective management. Within our knowledge rich approach, the management of

those artifacts is achieved collaboratively and requires reference models for comparison e.g. ontologies. Our work in the first phase of the project is as follows:

Method: Knowledge rich methods rely on analysis of the text by using knowledge sources outside the text (e.g. linguistic and domain ontologies, lexicon, dictionaries, grammars, etc.) for reasoning about the semantics (e.g. similarity in text meaning) and supporting sentiment analysis (Moilanen and Pulman, 2007; Liu, 2008). An annotation grammar is developed to mark concepts in the text. Ambiguities are resolved by means of discourse segmentation and lexical chain analysis, and the author's attitude toward the concept is determined through sentiment analysis. The information obtained through this processing chain is included in the vector space model of the knowledge poor approach in proper balance to obtain optimal text classification used for positioning.

Results: The CLaRK system (Simov et al., 2001) has been adapted to accommodate the processing chain including an annotation grammar, discourse segmentation, lexical chain analysis and sentiment analysis. As the process of developing the lexical semantic resources is data oriented and we need to have already available the data sets to be used in the validation available to be able to build such resources. We have already completed the analysis of data requirements for validation of WP 4.1 scenarios and the data set is being built.

Conclusions and further work: We will evaluate our knowledge rich approach for positioning by creating manually a gold standard corpus and then test our approach by means of precision and recall metrics. In addition we will compare this approach with the LSA-based approach with the aim to find an optimal combination of both in order to satisfy task goals.

Work package 4.2: Conceptual development

In WP4.2 we build on the work of Stahl (2006) to provide a tool to support the development of the individual learner, providing a component of an individual's reflective learning cycle and corresponding to the 'building personal knowing' in the Stahl cycle.

Method: In the first phase, we compared the outputs from a number of concept mapping tools providing a means of determining how learners relate basic concepts, to establish which could meet the requirements identified in the WP4.2 scenario. We investigated the utility of reference models against which learner texts can be compared, as a basis for feedback.

Results: Although Leximancer and Pathfinder were selected for initial experiments, their functionality and flexibility was insufficient for the requirements of the project. We will therefore develop a custom tool based on LSA as a basis for the WP4.2 service. Using these concept mapping tools, a clear distinction could be made between the ability to integrate concepts demonstrated by individual learners with that of reference material. Two types of reference models were investigated: a "pre-defined reference model" based on materials from the curriculum and an "emerging reference model", drawn from the concepts and inter relations between them generated by a peer learning group. The pre-defined reference model was too complex for comparison with a novice learner and may not be suited to early stages in a curriculum. The emerging reference model was a better indicator of the appropriate level of abstraction and relationship between concepts attainable by individual learners.

Conclusions and further work: We conclude that the emerging reference model, based on the relationships between concepts, generated by peer groups of learners, may be more useful to the novice learner as it approximates to his/her Zone of Proximal Development. Future

studies will investigate the effectiveness of comparisons of the pre-defined and emerging reference models for different groups of learners. Stakeholder feedback will inform the development of the custom LSA-based service to provide formative feedback.

1. Introduction

1.1 Purpose of document

In this report we claim that learner positioning and diagnosis of conceptual development can be addressed by applying the latest advances in the research domain of Natural Language Processing (NLP), and particularly Latent Semantic Analysis (LSA). To this end, this report attempts to answer the following questions:

- Which educational theories are relevant to develop a solution to tackle these problems?
- How will these theories relate to the proposed solutions for these problems?
- What are the solutions that will be developed?
- What technologies and methods will be used or developed to implement the solutions?

1.2 Project goals

The Language Technologies for Lifelong Learning (LTfLL) project is concerned with adult learning in the context of the new language processing technologies and of the collaboration facilities offered by the Web2.0. Adult learners usually bring with them significant prior experience and will show a degree of autonomy in their learning (Knowles, 1975). Although this was emphasised in adult learning theory, subsequent research has shown that it is relatively inefficient and ineffective for adult learners to be wholly self-directed (Dornan and David, 2000). Their motivation is increased by external sources of support such as feedback, peer comparison, and mentoring (Sargeant et al., 2006). The success of the new paradigm of Computer-Supported Collaborative Learning (CSCL, see Stahl, 2006), as exemplified by Web2.0 tools and social-cultural learning theories (Vygotsky, 1978), adds a collaborative, social dimension to classical, autonomous learning.

LTfLL focuses on adult learning taking place in the work place, in vocational studies and in Higher Education programs. The aim of LTfLL is to create services that enhance individual and collaborative development of competence in educational and organizational settings. The intention is to use language technologies extensively, so that adult learners can be supported effectively and efficiently.

Learner support can place a heavy load on staff time and resources. Stakeholder analysis has identified four types of activity that are responsible for this burden: assessment of student contributions, answering students' questions, community and group support and monitoring and assessing the progress of students' studies (Van Rosmalen et al., 2008). The development of services in work package 4 addresses these issues.

1.3 Tasks of work package 4 and their relation to other work packages

Work package 4 (WP4) is dedicated to developing the means of positioning the learner, through monitoring and assessing study progress. Its particular aim is to support lifelong learners as individuals:

- by assessing what they know in order to recommend appropriate subsequent learning materials (task WP4.1)
- by helping learners recognize their understanding of a particular topic, so that they can develop further as independent learners (task WP4.2).

Work package 4 uses the scenarios developed in WP3 in association with validation activities (WP7) to guide the design of its services. Working with WP2, the services will be embedded in the LTfLL Personal Learning Environment and where appropriate, will interoperate with services in other work packages. The outputs from WP4 will comprise positioning services based on knowledge-poor and knowledge-rich approaches, in association with user interfaces delivering the required functionality to learners, tutors and other stakeholders. As well as establishing how language technologies can best be used to provide the underpinning positioning and conceptual development services, an important challenge (given the complexity of data output from the underlying services) will be to establish effective ways to deliver useful information to end users.

As stated in the LTfLL Description of Work, Medicine and Information Technology have been chosen to explore the ideas and services to be developed. Both domains include learning in formal and informal settings and learning, which may lead to certification. In WP4, the undergraduate Medicine programme provides a good model for self-directed learning in the work place as well as for more traditional learning in Higher Education. Medicine is also of interest for its multi-disciplinary nature. Practitioners are expected to adopt a holistic approach integrating an understanding of disease processes with communications skills, psychology, sociology and the ethical implications of healthcare, underpinned by an understanding of how to learn. In contrast, our commercial partner BIT-MEDIA provides short courses in Information Technology to unemployed adults, where the emphasis is on assessment of knowledge and skills.

WP4 is concerned with textual evidence of learning, which is an important medium for communicating knowledge in education and through electronic media. Although some learners use other media to represent their learning, older and more recent socio-cultural perspectives see language as a key mediator of learning (Vygotsky, 1978, Wertsch, 1991). According to socio-cultural conceptualizations, construction of knowledge through dialectical, dialogical and social processes results in a large reservoir of tacit knowledge (See Figure 1). In addition, the social cultural interactions have effect not only in shared, tacit knowledge but also in linguistic patterns of usage (speech genres; Bakhtin, 1986). Analysis of text is therefore a valid means of achieving the aims of this work package.

1.4 Positioning and language technologies (WP4.1)

The purpose of WP4.1 is to provide the underpinning technology for the project to support positioning using language technologies. WP4.1 has worked in collaboration with WP3 to provide a real life scenario, which will be implemented and will demonstrate the use of advanced positioning technologies in real life situations.

Learners, coming from a variety of backgrounds, have different learning goals and different prior knowledge. Learners' knowledge gained from previous experiences of learning can be of a formal nature (certified exams, certificates, etc.), in which case standard

admission/exemption procedures may apply, or non-formal learning where such standard procedures are not available. Positioning in such contexts refers to the identification of a learner's existing knowledge and to the comparison with knowledge existing within relevant Communities of Practice (Wenger, 1998). In order to provide relevant text-based knowledge resources to learners, enabling them to best direct their learning efforts, we will explore whether language technologies can be used to recommend to learners the most appropriate text-based knowledge resources in relation to their current position.

A key research problem in order to operationalise the scenario is to establish the best ways in which to optimise the use of language technologies to achieve meaningful information for the stakeholders, e.g. learner, tutor. We will study samples of real life texts in the medical and IT domains in order to establish answers to the following questions:

- *To what extent can Language Technologies provide a mechanism to analyse and compare evidence of previously acquired knowledge that identifies the learner's position in a given domain?*
- *To what extent can these language use patterns be used as formative feedback for the learners?*

This analysis does not presuppose that language technologies can be used to examine a text and directly determining from that text what the author knows and does not know. As is discussed in Section 5, the determination of learning knowledge is indirect in the sense that it uses evidence based on phrases, terminology and general language use patterns.

This analysis does not presuppose that language technologies can be used as an alternative to tutor advice but rather as a support to it. That is of particular interest when these patterns are unconsciously used and would not otherwise be noticed by the tutor without the help of language technologies.

Our proposed solution is to start with a knowledge poor approach using LSA combined with a novel bag-phrase approach to compare learner texts with expert texts. Learners will receive quantitative feedback indicating the distance between these texts. In addition, learners will receive qualitative feedback indicating the fit of their language usage in relation to language used within relevant communities of practice. As second step, we will incorporate a knowledge rich based analysis to provide similar feedback.

1.5 Conceptual development of the learner (WP4.2)

The purpose of WP4.2 is to provide the underpinning technology for determining the conceptual development of the learner as the basis for providing feedback to individuals and facilitators, e.g. tutors. WP4.2 has worked in collaboration with WP3 to provide a real life scenario in the Medicine domain. The scenario will form the basis for a conceptual development service which will be tested in real life situations in order to provide iterative improvements to the conceptual development engine and associated functionality, e.g. visualisations and reports.

Formative feedback aims to communicate information that will engender accurate, targeted conceptualizations of a particular topic for the purpose of improving learners' understanding

of it (Shute, 2008). Learners, therefore, need to monitor their progress in their understanding of a specific area or problem, so that they recognize the limitations of their current level of expertise. Both these limitations and the degree of progress made will determine the level and type of formative feedback they need. Monitoring and recognizing a learner's level of expertise has to take into account their knowledge, the level of cognitive processing required by the task, and the associated instructional strategy used (Ertmer & Newby, 1993; Jonassen et al., 1993b).

The key research problems are to establish the best ways to determine conceptual development and the most effective means of communicating the results to the stakeholder, e.g. learner, tutor. We will work closely with stakeholder groups to determine the answer to the second problem, through validation of WP4.2, in collaboration with WP7. Our overall questions are:

- *To what extent can Language Technologies provide a basis of formative feedback service for individual learners, which takes into account their conceptualization of a topic?*
- *How does this approach with compare other methods and tools that can provide such formative feedback?*
- *Can it be used to best advantage if it is in combined with them?*

Our proposed solution is to build an LSA-based service that provides a comparison of the LSA analysis of learner texts with reference models, to provide meaningful aggregated information to stakeholders (learners, tutors).

1.6 Relationship between WP4.1 and WP4.2

In summary, both WP4 tasks rely on information extracted from texts. They differ in that WP4.2 extracts concept-like clusters from the texts, whereas WP4.1 relies upon surface level phrases and patterns of usage. Workpackage 4.2 is in this sense “knowledge rich” since the concepts are knowledge-like units extracted from the texts. For WP4.1 there is also knowledge rich subtask (Section 5.3) relying upon concepts but within this subtask the concepts are determined externally by use of ontologies. Thus, the knowledge rich subtask of WP4.1 provides a bridge between the knowledge poor subtask and WP4.2, as they are both using notions of “concepts”, which will be defined more precisely in the respective sections.

The two tasks also share a challenge in working out how best to aggregate highly complex underlying data in ways meaningful to end users to meet their business needs. This will be addressed through validations of the user interface and associated reports with end users, to enable fine tuning of the outputs to end . This will be achieved in collaboration with WP7 and according to the results obtained, enhancements to the scenarios may be indicated in collaboration with WP3.

1.7 Overview of this report

This document is based on work taking place up to December 2008 ('first phase'), plus remedial work required by European Commission reviewers. Later work will be reported in deliverable D4.2, due in Month 20. This document should be read in association with the

separate deliverable "LTfLL consortium's approach to integration – additional report" (September 2009), which explores the integration of WP4 services with those of WPs 5 & 6.

Section 2 of this report discusses the theoretical background to adult learning and how our work relates to these theories.

Section 3 describes the educational domains in which the work of WP4 is situated.

Section 4 describes the texts from these domains that are available for analysis.

Section 5 describes the knowledge poor and knowledge rich positioning services, the work during the first phase, the conclusions drawn and the proposed next steps.

Section 6 describes the service for diagnosis of conceptual development, the work during the first phase, the conclusions drawn and the proposed next steps.

Section 7 summarises progress to date and draws overall conclusions.

2. Theoretical background

2.1 Introduction

We will now set this work package in the context of relevant developments in educational theory, perceptions of how knowledge and understanding are acquired and of learner development. Adult learning in a work place environment is recognized as a social activity, so we will discuss learner development from this perspective. We will then consider how this relates to the development of the individual learner and discuss its application to Medicine.

2.2 Theoretical basis of learning

It is extremely important to state clearly the theoretical perspectives of learning considered in LTfLL, their limitations and the possibility of integrating them, in order to assure coherence among all the work-packages in the project. Epistemological questions about education theory are concerned with the relationship between “the knower” and what can be known, and the extent to which learning concerns ‘real’ external entities, or social constructed truths. Learning theories that are concerned with learning as absolute truth that can be understood through an ‘objective detachment’ are characterized by Guba & Lincoln (2005) as ‘naive realism’, characteristic of a ‘positivist’ system of beliefs. In contrast, a ‘constructivist’ system of beliefs is focused not on knowledge as an absolute external reality, but on its construction by the knower, individually and as a member of a social community.

There has been a progressive shift towards a more constructivist epistemology of education in recent years. Furthermore, Wertsch (1991), continuing Vygotsky's cultural-historical ideas (Vygotsky, 1978), has discussed the need for socio-cultural perspectives on learning in addition to traditional psychology, on the grounds that the latter tended to study the individual “in vacuo”. Moreover, Wertsch (1991) has emphasized that a Vygotsky's ideas are extended by Bakhtin's dialogism (1981). Bakhtin even considers that every text, not just conversations, are dialogs in which multiple voices interact. This idea is very important because it can provide a theoretical foundation for social knowledge construction. Mikhail Bakhtin (1981, 1986) considers that dialogism is not limited to conversation but it is rather a general phenomenon that occurs even in written “utterances”. Always there are several voices that interact, for example, the writer, the potential reader, the echoes of the voices present in each word. Moreover, from this multivocality perspective, texts become meaning generation mechanisms, facilitating understanding and creative thought, as Lotman stated (Wertsch, 1991; Dysthe, 1996). A consequence is that in education, “the interaction of oral and written discourse increased dialogicality and multivoicesness and therefore provided more chances for students to learn than did talking or writing alone” (Dysthe, 1996). The dialogic and multivoicesness features of any utterance, even written, may be unifying factors for the integration of the modules in the language-centered LTfLL project. Therefore, an integrated framework is provided for analysing all the textual learning activities as searching documents, reading, writing summaries or forum posts and chatting.

For LTfLL, the two dominant theoretical perspectives are cognitive and social. A *cognitive* orientation focuses on perception, memory and meaning; it assumes the memory is an active processor of information, and knowledge, as a commodity plays an important role in learning.

A *social* orientation assumes that learning is a social activity, which occurs in interaction with others. It takes account of both the learner and the environment, where learners are not just products of their experiences, but pro-active producers of the environment in which they operate. Socio-cultural learning theory, of particular relevance to the workplace learning environment of this project, originated in the work of the Russian scholar Vygotsky and had its roots firmly in Marxist notions of collaboration (Wertsch, 1991). Its emphasis is on learning as an essentially collaborative activity, where not just the processes but also the products of learning reside in ‘activity systems’ or ‘communities of practice’ (Lave and Wenger, 1991). Socio-cultural theorists are concerned with learners’ social engagement in communal activity and the identities, language, and cultural artefacts of the social groups in which they learn.

2.3 Knowledge creation theories

Knowledge creation theories focus on how individuals and groups develop knowledge that is new to them. They stress that knowledge is not transmitted untouched and unchanged from one – knowledgeable – person to another person who is unknowing. In contrast, they emphasize that knowledge is constructed in a dialectical and social process, and that not only explicitly stated knowledge and information is a source or result of this process but that there is also a much bigger reservoir of tacit knowledge (Figure 1).

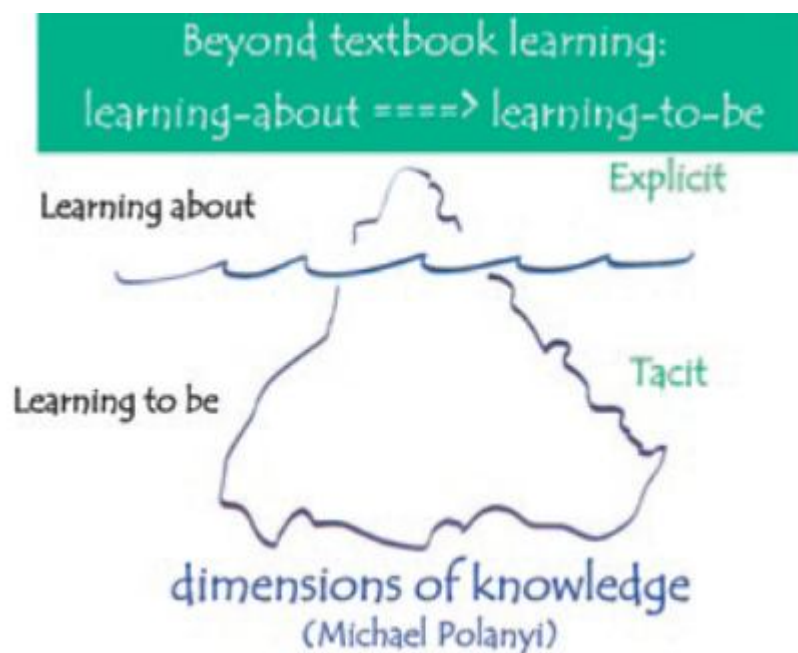


Figure 1: Explicit and implicit knowledge (Brown & Adler 2008)

Contemporary trends in educational and organizational contexts share this view of how knowledge is created. As described in the integrated report, in the area of collaborative learning, Stahl (2006), following a social epistemological perspective (Brown & Duguid, 1991; Lave & Wenger, 1991) models the learning process as a mutual construction of individual and social knowledge building.

Figure 2 shows Stahl's cycles of knowledge building. The diagram depicts the interaction between the cycles of personal and collaborative knowing building. The lower left corner shows the cycle of personal understanding, which can start with a tacit pre-understanding influenced by personal knowing. The right part of the diagram depicts how the social process of interaction with people and with our shared culture influences the individual's understanding. Although in the diagram personal understanding and social knowledge building are separated, it is only a matter of representation, they can only be separated artificially. Our motivation for introducing the Stahl cycle is that it fits both WP4.1 and WP4.2 approaches. Moreover, as it was described in the integrated report, WP5 and WP6 also fit this model.

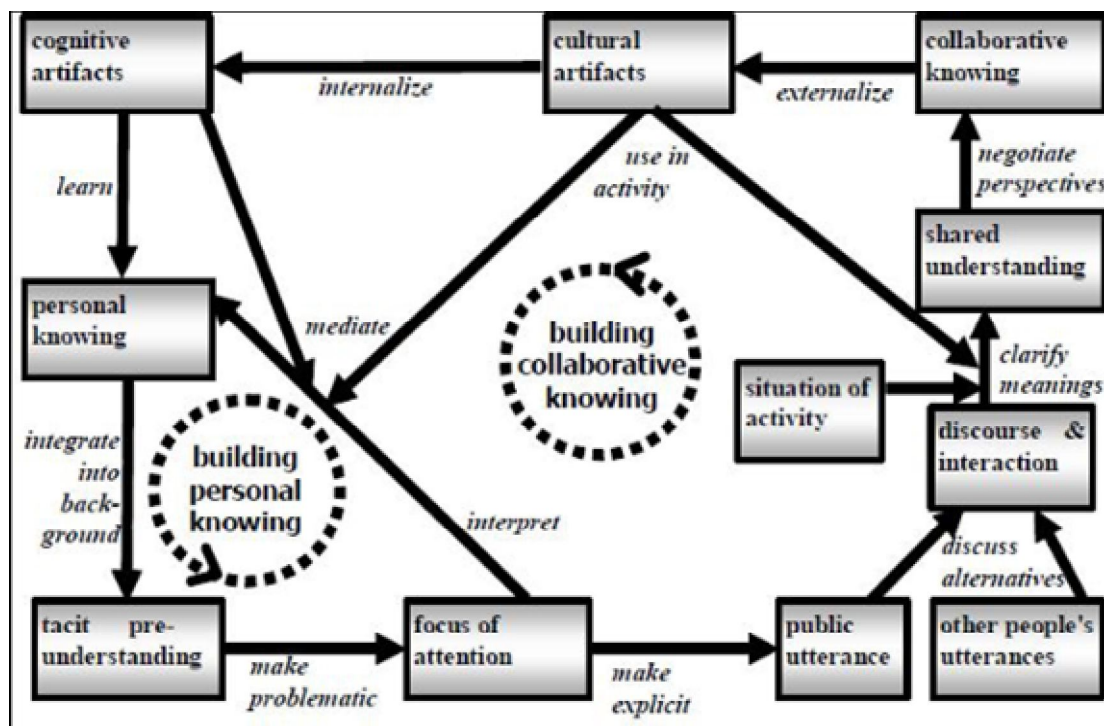


Figure 2: Cycles of knowledge building (Stahl 2006)

Cycle of personal understanding. As indicated in Figure 2, learning may begin with tacit pre-understanding, which may change if we clarify the implications of that understanding and resolve conflicts in our perceptions or fill gaps — by reinterpreting the basis for our knowledge — in order to arrive at a new understanding. This typically involves some feedback: from our experience with artefacts such as our tools and symbolic representations. It is noteworthy that this parallels the constructivist view of “self” development, described in Kelly’s Personal Construct Theory (Kelly, 1955), where man is viewed as “scientist” constantly testing and retesting his experiences of the environment in order to construct a view of himself as an individual. This is discussed further in relation to the development of learner identity in medicine. This new comprehension is embedded to become our new tacit understanding and to provide the starting point for and further learning. If we cannot resolve the problematic character of our personal understanding alone, then we might need to enter into an explicitly social process and create new meanings collaboratively. To do so, we typically articulate our initial belief in words and express ourselves in public statements,

entering into the cycle of social knowledge building. Conversely, our problems may arise thorough social interaction and need to be solved by the individual.

Cycle of social knowledge building. (Building collaborative knowing) In this cycle the interchange of arguments from different perspectives, may converge on a shared understanding if differences in interpretation and terminology are clarified through negotiation, thereby culminating in an accepted knowledge. This process can therefore be viewed as a continuum between personal understanding and generally accepted knowledge.

The knowledge poor subtask of WP4.1, with its focus on surface language, is represented mostly by the right hand side of the cycle since language is the mean of social communication. It is understood here that experts in a given field develop a speech genre (Bakhtin, 1986), which includes terminology and phrasal usage. Thus, a learners degree of expertise can be indirectly measured by textual distance to expert usage. Moreover, feedback concerning phrases and usage can facilitate communication, and ease the learner's integration into the community of practice, leading in turn to increased social learning.

WP4.2 is represented in both cycles. In the left hand side of the cycle, it provides a cognitive artifact (i.e., individual visualization of learner's textual inputs) that can help learners to understand and resolve conflicts or filling in gaps of their knowledge. If this is not possible, learners enter into the cycle of social knowledge building. In this cycle, WP4.2 provides a 'cultural artifact' (i.e. an amalgamated visualization of all peers textual inputs), which can be seen the joint understanding of that group (at that moment in time) that can help to foster shared understanding. This is of particular importance to WP 4.2 as a reference point for an individual learner can be the concepts and the relations between those concepts that a group of people (e.g. peers, participants, co-workers, etc.) used most often.

It is important to note that from a cognitive viewpoint, the Stahl's diagram does not represent the skills and sub-processes required for learner development in this context, for example personal skills, like summarizing discussion, understanding, texts, critical thinking and logical structuring of arguments; social interaction skills such as turn-taking, repair of misunderstandings, rhetorical persuasion and interactive arguing. Of particular significance, this also includes activities for providing feedback in both cycles, to support personal understanding and social knowledge building.

The interplay between the individual and group learning has also been described by, Nonaka, Toyama, and Konno (2000) who identify four connected and interacting processes of knowledge conversion, together the "SECI"-process:

- **Socialization** - the permeation of tacit knowledge through and between groups through shared experiences
- **Externalization** – the articulation of implicit knowledge through distinguishing phenomena and episodes. Conceptualization and mental modelling are the basis for two processes: combination and internalization (see below)
- **Combination** – where explicit knowledge is critically analysed and combined with other explicit knowledge or restructured into more complex new knowledge
- **Internalization:** the process in which new explicit knowledge is embodied, connected to new contexts and made useful and productive; the implicit knowledge that is then accumulated can start a new SECI-cycle (Figure 3)

This can take place at different levels of sophistication, depending on how people create and employ a context for implicit and explicit communication, the quality of the input in the process, etc.

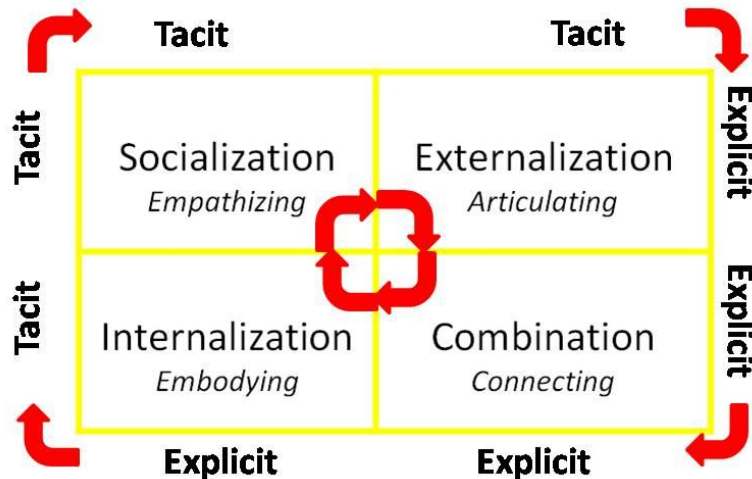


Figure 3: SECI spiral and subprocesses (Nonaka, Toyama & Konno 2000)

Many educational practices start by providing students with explicit knowledge, and only after this has reached what is considered a critical mass, are they allowed to acquire implicit, experiential, applied knowledge. Ertmer and Newby (1993) and Jonassen et al. (1993b), however, do not advocate to a single theory of learning, but emphasize that the instructional strategy and the content addressed depend on the level of the learners. They claim, therefore, that behavioural strategies can facilitate mastery of the content of a profession (knowing what); cognitive strategies are useful for procedural knowledge (knowing how); and constructivist strategies are appropriated to dealing with ill-defined problems as summarized in Figure 4

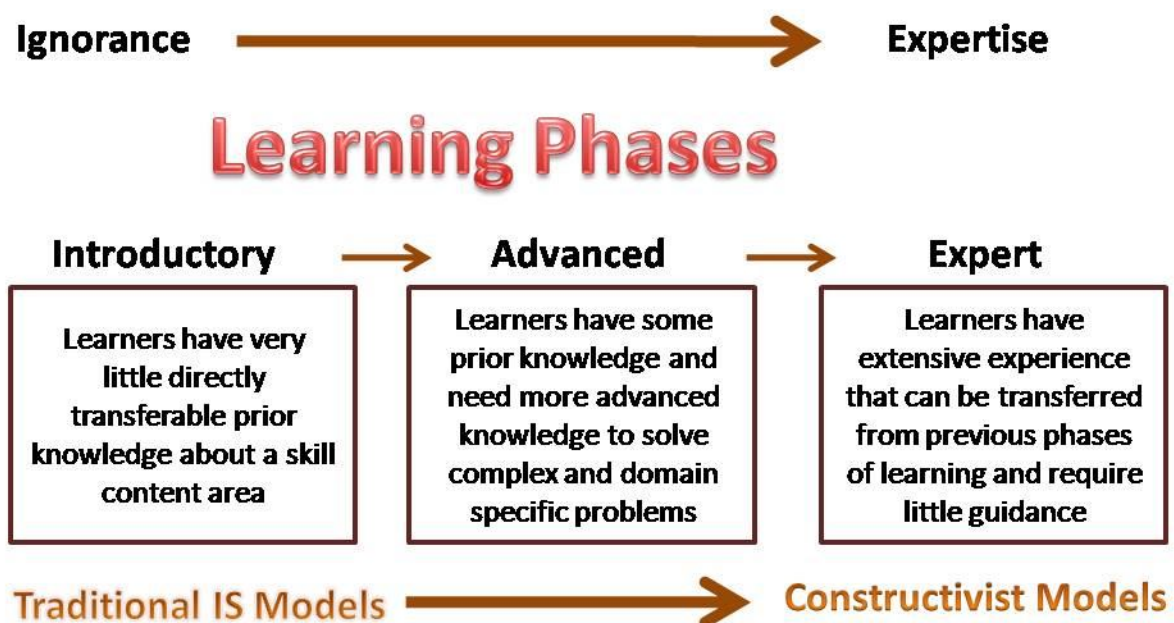


Figure 4: The Continuum of Knowledge Acquisition Model (Jonassen et al., 1993b)

Ertmer and Newby (1993) believe that instructional strategies depend on the level of learners' task knowledge and the level of cognitive processing required by the task. From our point of view, this implies as well that while monitoring and providing formative feedback (Section 6), the level of expertise of the learner should be considered.

2.4 Theories which model the gaining of expertise and their relevance to professional development

Theories which explain and predict how learners develop expertise in a specific domain involve growth paths, an understanding of the essence of expertise, strategies used by learners and teachers/tutors ("instruction strategies) and physical and/or cognitive changes that occur during this development. One approach is described by Ericsson (1996, 2004) in the theory of Deliberate Practice (DP), in which the key concept is that practice in a particular field must be informed by a good analysis of the present state of mastery, which targeted at the improvement of specific points, with the help of well-chosen teachers, models, and other support persons. Ericsson also showed that different stages of development require different teaching strategies and that the best predictor of the final level of expertise reached was the accumulated time spent in deliberate practice. The key criterion for defining expertise is superior performance. This is theoretical approach is extended to professional domains (Ericsson, 2004). Others (Mieg 2009, Sternberg and Frensch, 1992), however, include the function of experts within their community of practice in their definition, termed the "attributional definition". In contrast, the Model of Domain Learning, MDL (Alexander, 2003), describes development of expertise in three increasingly advanced levels (stages): Acclimation, Competency, and Proficiency. Within these stages, three interrelated dimensions are proposed that change with level of expertise: (i) domain knowledge that undergoes both quantitative and qualitative changes, (ii) learning strategies that are related to depth of knowledge of a domain, and (iii) interest that varies along axes of generality, and context-dependence.

Models for acquiring professional expertise are positioned in between the MDL and DP theories but are also significantly different from them in some key aspects (Boshuizen and Schmidt 2008). They focus on the learners' commitment to specific domains and on the learning environments, both in an academic and workplace situation. They consider requirements for acquiring competences, integrating knowledge, skill and developing professional attitudes. This is further discussed in relation to Medicine. DP may not be completely applicable in these situations, even if good or excellent performance can be discriminated from those which are mediocre, it can be difficult to reach a consistent elite level, due to the breadth of the domain (different but interrelated specialities in medicine), or continuous changes in the environment or in the subject of the profession (e.g., investment, meteorology). In studies of professional expertise development, the definition of expertise levels is less strict, and experience and reputation are used as proxies. Although the models are tightly linked to the professions, they are able to make general predictions and, like MDL and DP, focus on knowledge structure and instruction strategies. This further emphasises the inter relationship between the basis of expertise development and the understanding of knowledge building and acquisition, discussed previously.

Expertise theories, more or less adhere to the definition that experts do a better job than non-experts (or that professionals do a better job than non-professionals). Some of them also state that experts are able to move their field further, by being well acquainted with key issues and being actively engaged in problem solving (Alexander 2003) Although Ericsson (1999) also states that high expertise and innovation are intricately related, prolonged time practising in a profession may not equip the professional to cope with a changing environment, and the individual may be prone to poor practice and skill-obsolescence (Weggeman, 2000; Thijssen & Van der Heijden, 2003), Expertise should, therefore also include flexibility and employability (Van der Heijden, 2000) and exhibit routine and adaptive expertise (Hatano & Inagaki, 1986) and emphasizes the need for training for transfer at every level of learning (Nokes & Ohlssen, 2003; Salomon & Perkins, 1989). This is an essential component of lifelong learning and is an important basis for WP 4.1.

2.5 Knowledge restructuring

In MDL (Alexander, 2003) the process of gaining proficiency involves an extension of knowledge and connecting between concepts. This process of knowledge restructuring has been intensively investigated in medical education and is especially important to re-evaluating how basic sciences knowledge is introduced and used. Thus, knowledge encapsulation, namely forming macro concepts (encapsulations) that subsume biomedical concepts under clinically relevant headings (e.g. in terms of diagnosis, patient management or treatment), plays a prominent role (Schmidt et al., 1990). Biomedical knowledge provides structure to isolated clinical case concepts (Woods, 2007; Woods, et al., 2005). Woods et al. (2007) demonstrated that learning causal explanations for features of clinical conditions resulted in ability to make a quicker and more accurate diagnosis in complex situations. Illness script formation (Schmidt et al., 1990) is another knowledge restructuring process reviewed in detail by Charlin et al., 2007. This process takes place under the influence of practical application of knowledge and is summarized in Table 1. Even in early stages of learning, application of knowledge is linked to the conditions in which this knowledge is acquired. This non-analytic aspect of knowledge application is rapid and not under conscious control. Depending on the context of learning and later application the effects can be advantageous or disadvantages (see Norman et al., 2006, pp 344-347, for an overview). Enabling learners to understand how they are restructuring their knowledge, in a manner that is appropriate both to their level of expertise and to the social context of their learning, is an essential feature of the services that Work package 4 aims to provide.

Table 1: Knowledge restructuring, clinical reasoning and levels of expertise (Boshuizen & Schmidt, 2008; reprinted with permission from Elsevier)

Expertise level	Knowledge representation	Knowledge acquisition and (re)structuring	Clinical reasoning	Control required in clinical reasoning	Demand on cognitive capacity	Clinical reasoning in action
Novice	Networks	Knowledge accretion and validation	Long chains of detailed reasoning steps through pre-	Active monitoring of each reasoning step	High	Difficulty to combine data collection and

Expertise level	Knowledge representation	Knowledge acquisition and (re)structuring	Clinical reasoning	Control required in clinical reasoning	Demand on cognitive capacity	Clinical reasoning in action
			encapsulated networks			evaluation and clinical reasoning
Intermediate	Networks	Encapsulation	Reasoning through encapsulated network	Active monitoring of each reasoning step	Medium	...
Expert	Illness scripts	Illness script formation	Illness script activation and instantiation	Monitoring of the level of script instantiation	Low	Adjust data collection to time available and to verification/falsification level of hypotheses

2.6 Professional learning

In the context of professional learning and development, Schön (1987) approached application of knowledge to professional problems and situations from the perspective of the practices in which learners were trained and in which they work. A key element is *reflection-in-action*, which results from monitoring one's own action and is at least partly conscious. It involves recognizing unexpected phenomena, and the consequences of "knowing-in-action" i.e. the active use of knowledge and understanding. Reflection-in-action can provoke questioning of in-action knowledge and may lead to a revised understanding and experiments with problem-solving strategies, etc. Although Schön's theory takes this in-action aspect for granted, practices and professions have different time structures, which may constrain the possibilities for reflection-in-action (Eraut, 1994), for example flying an aircraft, or operating on a patient. Eraut suggests high-speed practices require instant recognition and response, and routinised, unreflective action at the time but would require subsequent reflection post action, for example in the medical audit process. Low speed practices allow deliberative analysis and decisions, with actions following a period of deliberation. The mode of cognition in intermediate situations would consist of rapid interpretation and decisions, and action monitored by reflection. Working in low-speed workplaces can go hand in hand with learning. Medium and high-speed work requires both preparation of learners, to help them become aware of what can be expected, and reflection-after-the-fact. The ability of a learner to understand his/her learning process and to respond to feedback provided by peers, tutors or experts is therefore essential to reflective practice

2.7 Learning processes

The learning processes taking place in education either in formal or non-formal learning contexts, both undergraduate and post graduate, are the same as those described for expertise development: knowledge accretion, restructuring and script development. Which of them presents most difficulty to students depends on the domain involved. Knowledge of these difficulties is part of Pedagogical Domain Knowledge, PDK (Shulman, 1986). Those in mathematics are different from those encountered in medicine or law. Conceptual change and the resistance to it is a problem shared by many domains (Vosniadou & Ortony, 1989). Approaches have been formulated which may help learners reach a change in their understanding of key concepts (Chinn & Brewer 1993), for example integrating several different perspectives of a specific problem (Spiro et al., 1992) or the development of specific problem-solving scripts (van Merriënboer et al., 2002). The teacher/tutor may be regarded as encapsulating PDK. They develop their approach to learner's difficulties in their daily interaction with them within their domain of study and use this knowledge to adapt instruction. A good teacher adapts instruction and feedback to the learner's Zone of Proximal Development (Vygotsky, 1978) and can be considered part of the teacher/tutor's PDK. Hattie's (1996) meta-analysis of the effectiveness of teaching strategies showed that the most powerful strategies are:

- Provide feed forward and feedback information that tells students how they are doing, where they are going, how they are going and whereto next; help and stimulate learning from feedback.
- Provide real challenges and communicate learning intentions and success criteria.
- Provide opportunities for modelling, both live (by expert and coping peers) and in the form of worked examples.

3. Educational domains included in WP4

3.1 Medicine as a model

Modern medicine requires clinicians to be independent life long learners, with a clear view of the limits of their expertise and competencies. The continuing advancement of research in medical and human sciences necessitates constant adaptation and revision of practice in the workplace to ensure patients' care and treatment are both safe and effective. The ability to sustain lifelong learning by doctors is best achieved by a self – awareness of how the individual learns. Instilling the characteristics and behaviours of independent lifelong learning in doctors is therefore major goal of modern medical education. Learning in the clinical workplace is typified by interactions and exchanges with fellow learners, senior clinicians, other healthcare professionals (nurses, pharmacists etc.) and, most importantly patients. They exemplify models of Communities of Practice (Lave and Wenger, 1991). In undergraduate medical education, approaches to learning are used which aim to accustom students to this environment, for example Problem Based Learning (O'Neill et al, 2002). Bleakley and Bligh (2008) describe modern medical education as patient focussed, identifying the learning of a young trainee doctor as a clear example of situated learning, focussing on information gained from patients by the learner's interaction with them. All these interactions represent learning experiences, on which the learner reflects and from which they construct their own knowledge, similar to that described in Stahl's model, which we have discussed previously. The introduction of the learner at a junior level into this environment also raised issues of their personal and professional development, even at an undergraduate level. Their perceptions of self identity as a learner, which we discussed previously in relation to Kelly's personal construct theory (Kelly, 1955) are significant influences on the importance that students attach to lifelong learning and reflective learning (Lown et al 2009).

3.2 Pedagogic domain: Reflective and critical learning in medicine – role of portfolio

The ability to think critically is key to enabling an individual to understand limitations of knowledge, competencies, skills, attitudes and behaviour throughout medicine but especially for medical students and newly qualified doctors. We have already discussed the basis for this in Schon's theory of reflection-in action and the requirement for post action reflection. An example is shown in Figure 5.

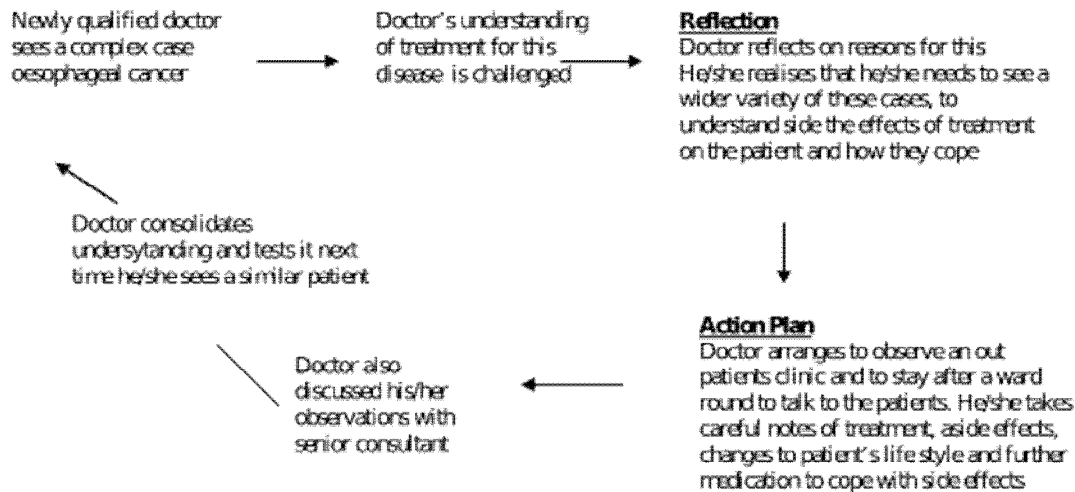


Figure 5: Example of reflection post action in Medicine

The awareness of such limitation is an essential prerequisite for obtaining feedback from tutors, seniors with expertise and acting upon it to formulate a learning plan for future development and is an important basis for WP4.

The evidence for that reflective learning has been implemented is now routinely maintained in learner's portfolios. These are widely incorporated into many professions; nursing, teaching, psychotherapy and counselling, law, engineering are all examples of careers which require mandatory (e)portfolios. The content of such portfolios are defined by the appropriate regulatory bodies for each profession and are required for scrutiny and evidence during regular professional appraisals. Its role has therefore developed from its earlier function as an indicator of an individual's learning and achievements to one in which the evidence that it contains clearly demonstrates the knowledge and competences required to progress within professions. The current view of portfolios (both electronic and otherwise) has changed since the DoW was written, and can no longer be regarded as a collection of essay type texts generated by the learner, but are more flexible and variable in content to accommodate the requirements of modern professional learning and development. Thus, the Centre for Recording Achievement (<http://www.recordingachievement.org/>) states that eportfolios should include criteria/standards/outcomes for learners (e.g. assessment); enable learners to create and store plans which can be shared with others (tutor/mentor/coach, possibly peers), which can be re accessed, and provide a means of recording experiences and achievements as they happen, which can be selected if they are relevant for a review by a senior. Learning plans and their re-assessment are relevant to the reflective learning cycle discussed above.

In the UK, all university students are expected to maintain a record of their personal and professional development, but also for medical students, the GMC stipulates that personal and professional development ePortfolios are an essential component of the undergraduate curriculum. They contain

- Evidence of engagement in activities that promote professional development (i.e. that is signed by the appropriate authorities, is dated and is verifiable)

- Evidence that the student is acquiring key clinical skills and the level of competency reached in those skills
- Critical self analysis of the student's own strengths and limitations and how they are building on the former and addressing the latter
- Evidence that the student is responding to the feedback given to him/her by those assessing development in skills and clinical practice
- Demonstration of how the individual student is developing their own learning strategies and clear indications that they can formulate their own learning plans, especially at key stages in the curriculum
- Evidence that they are developing an understanding of disease processes, which they apply to their clinical learning situations
- Clear indications that students understand and develop competencies in key features of modern medical practice e.g, team working, communications, governance of medical practice and evidence based medicine
- Critical thinking and reflective learning, which underpins many of the above features of the ePortfolio.

In summary, the “ideal” ePortfolio for the medical student is one which genuinely helps and supports the student's own personal and professional development. It demonstrates his/her development as a learner, both in terms of understanding and acquisition of skills and competencies, based on ever improving critical and reflective thinking and learning, thereby preparing the student for their life as a medical practitioner. In workpackage 4, the proposed automatic analysis (i.e. knowledge poor and knowledge rich based approaches), technology has the capability to inform and guide students through composite tasks of the “ideal” ePortfolio. The provision of feedback, supported by the services developed in WP4.2, and the learner's responses to this, would form important components of such portfolios. Conversely, some of the text material included in ePortfolios, for example participation on reflective online discussions, can be used as material for analysis with LSA. (See Section 4.2.2 below).

3.3 The role of ePortfolios in Computer Science education

One objective of the project is to apply the techniques and services, which we will develop, to more than one domain and language. We decided to choose computer science as the second domain for the following reasons: a) as one outcome of the LT4EL project (www.let.uu.nl/lt4el/), we have access to a large number of learning objects from the domain of computer science and in various languages; in addition we have access to lexica which have been derived from these learning objects and to an ontology connecting all the terms through language-independent concepts; b) some of the partners have experience in teaching computer science.

The situation with ePortfolios for this domain is far from ideal. In contrast to e.g. medicine, the use of ePortfolios is not an established practice in science and engineering (cf. Bhattacharya *et al.*, 2006). Within the consortium there are no partners actively using ePortfolios in Computer Science as part of their teaching. We will therefore generate eportfolio materials. In any case we are convinced that the practice of using ePortfolios will become more widespread, and therefore we see an increasing number of users of our service, which uses ePortfolios in order to position learners and monitoring their conceptual development.

4. Texts available for analysis

4.1 Introduction

Learner text that have been made available by the University of Manchester, for testing the tools and services that are being developed. The learner population is the undergraduate students of the University of Manchester Medical School. For WP4.1, text material generated in the curriculum by online discussion will be used, whereas for W4.2, conceptual development, a variety of text material will be used including learning diaries (blogs) created by individual learners specifically relating to problem based learning cases and clinical problems encountered in the work place environment and transcripts of "think aloud" protocol, described in section 6.5.1.1.

4.2 Text material generated in Manchester University medical curriculum

4.2.1 Theoretical Basis of Online Discussions

For undergraduate students in the University of Manchester Medical School, development of independent learning strategies and self directed study are necessary for progress in the curriculum, in which problem-based learning is an essential component. This is also reflected by the use of online learning, which is becoming more important in undergraduate medical education. It involves the learner as an individual but also can re recreate some aspects of Communities of Practice by using structured group discussions, to promote reflective learning and critical thinking. The contribution of individual group members to these discussions are therefore evidence of their interactions with others group members and of their own cognitive development (Braidman et al., 2008) The asynchronous nature of these discussions, allows time for reflection on workplace experiences, so that participants respond reflectively and critically in the context of their own experiences (Newman et al., 1997).

Learner and group development in these situations is explained by a Community of Inquiry model (Garrison et al., 2000), based on Wenger's Community of Practice concept, and reminiscent of Stahl's model, discussed previously. It explains development of such online groups as reflective learners through interplay between cognitive development, social interaction and influence of tutor/facilitator

4.2.2 Online Discussions in the University of Manchester Medical School

Asynchronous online discussions were introduced into Years 3 and 4 of the medical curriculum, when students are in a clinical workplace learning environment, with the purpose of enabling them to discuss issues concerned with personal and professional development. Large student numbers and their geographic dispersal required the introduction of electronically based discussion fora.. Students work in peer facilitated groups to discuss specific contemporary issues throughout the curriculum, with a nominated student taking on a facilitator role. There are 63 discussion groups in each of years 3 and 4, consisting of 8 students, including the student facilitator. All groups discuss the same topic, which changes each semester. Each text flow usually consists of at least 8 – 10 postings. Although some of

these are of just one or two lines, mostly they consist of at least 10 lines of text. All discussion groups participate i.e. the entire student population of Year 3 (452 students) participated and each discussion topic stimulates over 2,000 postings. Participation in these discussions is an essential component of the students' eportfolio

These discussion fora provide ideal datasets for LSA, both for comparative analysis and for corpus based analysis. For learner positioning, this provides some interesting possibilities across the groups, to suggest concepts that other student groups may explore, for example. It may also be possible to determine the progress of individuals within a particular group. An example is provided by discussions on the safe prescribing of medicines. This topic was initiated by suggestions to the students of the type of knowledge-based material on which to base their discussions, including some detailed compendia of drugs and treatment e.g. the British National Formulary. This may provide a sound corpus on which student discussions can be analysed and feedback provided to ensure good coverage of the subject area.

4.2.3 Data sets for LSA based positioning

The specific texts found most suitable as data sets for LSA were the online discussions related to the safe prescribing of medicines. These are based on a series of clearly defined learning objectives, which formed the basis of six main topic areas, which could be used for text analysis:

- The role of medical students in patient safety
- The role of the ward pharmacist in preventing serious medication errors
- The critical points in medicines management where serious errors can occur
- The swiss cheese model of accident causation
- The minimum core knowledge, skills and attitudes required to prescribe safely
- The legal consequences of negligent or reckless prescribing / administration of drugs.

Individual students' contributions have been selectively sampled based on a set of criteria that will provide a good quantity of content. Each contribution is presented in a format that includes the student identifier, the topic area that is covered and a grade (i.e. Excellent, Good, Fair, Poor) indicating the effectiveness with which the student has covered the domain topic.

4.2.4 Data generated by “think aloud” analysis of PBL cases

As discussed before, there is a major need to provide structured formative feedback to the students as individuals, of their conceptual development, resulting from their learning in their PBL sessions. An appropriate means of achieving this is to record a student's brief analysis of a recent PBL case and its relevance to the clinical condition concerned. The transcribed text output is suitable for language analysis that will allow a grading on which feedback to the student can be based. This is referred to as a "think aloud" protocol. The feedback will focus on their understanding of the subject area and their ability to relate it to the clinical problem originally posed. This method of text generation is especially appropriate to the medical domain as clinicians routinely provide verbal reports of clinical problems to colleagues, which often include a range of healthcare professionals.

5. Positioning the learner (Task 4.1)

5.1 Introduction

As life long learners have neither common learning goals nor common educational backgrounds, as it is the case in traditional learning settings, life long learning educational providers need to rely on available written materials produced by individual learners to identify their degree of expertise within areas of knowledge that are relevant to study programs that are offered. Lifelong learners seeking to further develop their competences will consider the different educational programs that best fit their needs. Accreditation or Recognition of Prior Learning (APL/RPL) provides a procedural solution for identifying the prior knowledge of learners in formal and informal education (Merrifield et al., 2000). This procedure is time consuming and maybe inaccurate therefore a (semi-)automatic procedure that addresses either or both of these two problems is worth considering. In this case, evidence of prior knowledge is provided by the learner in the form of texts, which need to be compared to public or domain knowledge. In this section, we describe knowledge-poor and knowledge-rich approaches to analysing learner texts. The WP4 description of work describes LSA and related methods (e.g. probabilistic LSA) that use text as the only source of knowledge as knowledge poor methods. Additionally, it describes the use of ontologies and lexical knowledge resources for the analysis of texts as knowledge rich methods. Knowledge poor based approaches can be considered as a base line for the knowledge rich approaches.

5.2 Knowledge poor based positioning

In this section, we introduce our knowledge poor approach to positioning. The approach that we will present is based on comparing words and phrases extracted from a learner's text(s) with terms and phrases extracted from model, expert texts. Overlap of these usage patterns will then be interpreted to mean that the learner has mastered parts of the expert domain. We will argue that this seemingly indirect approach is, in fact, reasonable. The more direct approach would be to analyze the learner's texts to determine what the learner knows and does not know. But this direct approach presupposes a degree of natural language understanding which is not available with current language technologies, and probably never will be available. Therefore, an indirect approach to assessing acquired knowledge is the only path available. We will argue that, since knowledge is, to a large extent, acquired socially, becoming an expert in a domain involves socializing with members of a community of practice, which in turn involves adoption of the speech genre of this CoP. Thus if a learner uses terms and phrases typical of this CoP, it is evidence that the learner may have socialized with the CoP, and has thus, most probably, also acquired some of the expertise of the CoP.

This line of reasoning is quite indirect, and clearly requires validation. We believe, however, that this is the only reasonable way to make sense of a knowledge poor based approach to positioning.

An added benefit of the approach is that it involves identification of words and phrases that are typical and atypical of a CoP. These phrases can be used to provide qualitative feedback to the user of the knowledge poor positioning service. If a learner uses phrases which are atypical of the CoP, the user can be informed of this inappropriate usage and more

appropriate, synonymous phrases can be suggested. If the user then adopts these more appropriate phrases into his or her linguistic usage, this will lead to facilitated communication with members of the CoP, which will, in turn, lead to increased social learning.

5.2.1 Integration of the positioning scenario within the LTfLL service platform

Although current scenarios for the positioning service implementation are relevant only to the Information Technology domain, the basic principles of our services to be validated by WUW and Bitmedia remains the same. Thus, this document focusses on the Medical domain.

Services providing support in determining the learner's degree of expertise or position by means of LSA in the IT domain will be implemented by WUW. The implementation of those scenarios is being guided by the philosophy of the relevant 4.1 scenarios written by Bitmedia and UTU. Bitmedia is a stake holder in the validation of the mentioned service and is currently working with UTU and WUW in developing the data sets required for that end (annotation of texts and training materials in the IT domain). A second service for positioning the learner will be implemented in a later stage using the knowledge rich approach.

5.2.2 Desiderata for (semi-)automatic positioning

We are interested in applying the techniques of text categorization for the purpose of positioning life long learners. Quite simply, we can rate learner texts, likely to be short and generated in informal educational settings, by using a vector-space comparison to gold-standard, expert texts. Then if the similarity is high enough, the learner will proficiency out of the course. This approach is straightforward, but in its naive version, it is unlikely to be successful. The problems concern accuracy, suitability and justification of the categorization. Accuracy is clearly important. False positives can result in learner frustration in courses beyond his or her level, and similarly false negatives can result in learner boredom. Suitability refers to model texts against which the learner texts are compared. If these are not suitable (or prototypical) models of expert language usage in the domain, then the positioning will not be valid. Thirdly, justification refers to the reasoning given by the service for the positioning decision. It is well known that decisions of expert systems are more accepted when the system gives reasons for its decisions. All three of these issues must be addressed by our system.

5.2.3 LSA based positioning

5.2.3.1 General introduction

In recent years, Latent Semantic Analysis (LSA; Landauer and Dumais, 1997) has been proposed as a suitable language technology for the automatic positioning of learners (Van Bruggen et al. 2004). Although, LSA has been successfully used in the context of language technologies enhanced learning (e.g. automatic assessment of student essays), learner positioning presents new challenges that expose the limitation of such an approach. In particular, learners produce text repositories containing few samples of text, many of them of small size and using language that is rich in non domain-specific expressions e.g. email messages, chat conversations, forum online discussions, blog postings, etc. In addition, those texts are generated in contexts where learners feel encouraged to hide their poor usage of language by articulating redundant expressions and making extensive use keywords. In addition, LSA is also limited in that it is not capable of recognizing directionality in causal relationships. Burek et al. (2007) presents a solution to this problem by means of a triple

based LSA that calculates a set of similarity measures between the semantically related constituents of the sentence structure (i.e. subject, verb and object).

5.2.3.2 Accuracy

The categorization obviously must be accurate. A false positive, indicating learner proficiency in a particular domain, could be dangerous as it could lead to work place incompetence. A false negative, on the other hand, could lead to boredom, as the learner is forced to take courses on topics that he or she has already mastered. The problem of accuracy is compounded by the fact that texts (selected from text collections are often short. To deal with this problem, our approach attempts to lose as little information from the text as possible. Traditional approaches to categorization lose information by case normalization, stemming and ignoring word order. The idea of the traditional approach is to deal with the data sparseness problem by collapsing textual features into equivalence classes, losing information in the process.

In our approach, we attempt to balance the problem of data sparseness with the goal of not losing information. This balance is obtained in two ways. First, we use LSA as a technique for dimensionality reduction. It is well known that LSA can be used to discover weighted clusters of words, which are loosely understood to be "concepts". Since these clusters can contain derivationally related terms, the need for stemming (and also case normalization) is reduced. Second, our more innovative contribution is to flexibly use n-grams of different size (phrases) and the above mentioned relational triples as opposed to strictly unigrams in the traditional bag-of-words model. Our approach to extracting such n-grams is to use an extension of the suffix array approach of (Yamamoto and Church, 2001).

5.2.3.3 Suitability

Suppose that learner texts could be accurately classified as similar or not similar to the gold standard text (or set of texts). Then the question arises as to whether or not the gold standard text is a suitable prototype for a good learner text. One approach to choosing a gold standard text would be to use a published journal article in the field. But such a text is unlikely to be similar to learner texts either in tone or in content. It is well known that effective teachers use scaffolding to present material within the zone of proximal development of the learner.

So perhaps a better gold standard would be a text written by an expert e.g. textbook, blog entry, etc. or other learning material, written at the level of the student. This is certainly an improvement, but on the other hand, it is still rather unreasonable to expect learners' texts to closely match the tone of a textbook, unless of course the learners are copying from the text. In fact, the texts that we have consist of online discussions of medical students on several topics related to safe prescribing. These texts have been categorized as to subtopic and graded for quality (excellent, good, fair, poor) by annotators at the University of Manchester. The texts contain serious conversations, with very little off-topic wandering. But the tone of the texts is chatty, and not at all similar to textbook writing. So rather than to use an external gold standard, we have opted for an internal gold standard. The prototypical "excellent" text is simply one that was rated as "excellent" by the annotators. But not all "excellent" texts are equally good as prototypes. Clearly, for any text t , the remaining texts can be ranked in order of similarity to t . If t is a good prototype, then this ranking should have other "excellent" texts as most similar to t and "poor" should be least similar. So we need to choose as prototypes, those texts that induce the best ranking. For purposes of comparing such rankings, we have experimented with the nonparametric permutation test.

5.2.3.4 Justification

As stated above, it is well known that users are more inclined to trust an expert system when the system can give some reasons for its judgement. Our system is designed to give both quantitative and qualitative feedback. The quantitative information indicates simply a distance measure of the learner text from the expert text. The qualitative information, on the other hand, uses the phrases that are extracted using the suffix array algorithms. These phrases are weighted according to the probability of occurring predominately in expert or non-expert texts. When the learner uses non-expert terminology, this can be reported to the learner as useful qualitative feedback. When the learner sees that the system is able to perceive subtle, perhaps unconscious, patterns of language usage, this will help to increase the learner's confidence in the quantitative judgement of the system.

The use of phrases, however, presupposes a fairly large body of training texts to overcome the data sparseness problem. Our phrase weighting approach is, however, flexible in the sense that if no distinctive pattern of use is detected for longer phrases, then the system will fall back to shorter phrases or single words.

5.2.4 Our progress to date

5.2.4.1 Phrases

Traditionally, text categorization by means of LSA has relied on a bag-of-words model. It seems, in some sense, obvious that a model based on phrases should be better. But it turns out that this is not necessarily the case. Recently, Bekkerman and Allan (2004) reviewed the literature on text categorization and found no general improvement when unigram models were replaced with bigram models. The problem is that using bigrams contributes heavily to the data sparseness problem. Bekkerman and Allan have, however, compared two rather extreme positions. Our idea is to extract phrases of any length from the training corpus, as long as the phrases are distinctive (occurring predominately in particular categories of documents). It may well be that the most distinctive phrases are generally phrases of length one (concurring with the bag-of-words model), but if there are phrases of other lengths that are more distinctive, then there seems to be no reason not to use these phrases. To give an idea of the approach, consider the word "side". In the medical discussions in our corpus, this word almost always occurs as part of the phrase "side effect(s)". In a few cases, "side" occurs in a unique context or as part of another phrase, such as "flip side". In this case, the distinctive phrase is apparently "side effect", and the other occurrences are just noise. These noise phrases are not only unhelpful for text categorization, they are also unhelpful for generating explanations that would be useful for learners and examiners. The example above raises some interesting counting issues. But first we need to specify more precisely what it means for a phrase to be distinctive.

5.2.4.1.1 Distinctiveness

In general, phrases that are evenly distributed across document categories are not very distinctive, whereas phrases that tend to cluster in one particular category are distinctive. This general principle must be applied carefully, however, since with small numbers, clustering may occur due to chance. A common measure of distinctiveness used for weighting in vector space models is tf-idf (term frequency multiplied by inverse document frequency) discussed by Salton (1988). It is unclear, however, that this is the best measure for picking out which

phrases to consider and which phrases to ignore. It is problematic, for example, that *idf* simply prefers terms that cluster in a small number of documents, regardless of the classifications. Given the ordinal classification of Manchester text set as “excellent”, “good”, “fair” and “poor”, we are not interested, for example, in terms that cluster in the “excellent” and “poor” texts. So a distinctive term should be one that occurs predominately in “excellent” and “good” texts or predominately in “fair” and “poor” texts. Consider, for example, the bullet point, with occurrence vector [31,5,0,0] [11]. The interpretation is that there are 31 occurrences in “excellent” documents, 5 occurrences in “good” documents and no occurrences in either of the poorer texts, this term appears be very distinctive of better texts. But if we count instead the number of different documents the bullet point occurs [4,1,0,0], we see a very different picture. The bullet point does occur in higher rated texts, but it is very bursty and is therefore not very useful for categorization. There are various approaches in the literature for dealing with burstiness. Since this is not our primary concern here, we deal with the problem by counting the number of texts containing a term rather than the total number of occurrences of the term. Thus, for the bullet point, we use the vector [4,1,0,0].

To rate a term such as the bullet point, we need some measure of goodness for the vector [4,1,0,0]. There is clearly no objective measure that can be used here. As a fairly reasonable score, we simply assign 1 point for every “excellent” text, 0.8 points for every “good” text and 0.2 points for every “fair” text. So, the bullet point receives a score of 4.8. This appears to be a good score, but what is the probability that a randomly chosen term appearing in 5 texts would have a higher or equally high score? We can answer this question by using a simulation. Random vectors are generated according to the known proportion of “excellent”, “good”, “fair” and “poor” texts. Then, for a high score such as 4.8, the idea is to count the proportion of randomly generated vectors have an equally high or higher score. And for a low score, the opposite idea is to count the proportion of randomly generated scores that are equal or lower.

5.2.4.1.2 Phrase extraction

In principle, the distinctness measure given above can be used with phrases of any length. If longer phrases can be found that are more distinct than single words, then there is no reason not to use the longer phrase. The problem is that the simulation-based distinctness test is very expensive, and it is certainly not possible to run this test for *n*-grams of every length in a text. The solution to this problem comes from Yamamoto and Church (2001), who show suffix arrays can be used to put the large number of *n*-grams into a much smaller number of equivalence classes. Using suffix arrays, it is very easy to pick out just the phrases that are repeated *n* times for some *n*, and it is very easy to extend phrases to the right: if “mumbo jumbo” repeatedly occurs together as a phrase, then it makes no sense to count “mumbo” by itself. Yamamoto and Church's suffix array program will put these two phrases into an equivalence class, so that that statistics can be calculated for the class as a whole rather than individually for all the members of the class. Since the time of Yamamoto and Church's paper, suffix arrays have been an active area of research, primarily in bioinformatics. One of the weaknesses of the suffix array approach used by Yamamoto and Church is that extensions to the left are difficult to discover. So it is difficult to discover, for example, that “jumbo” always combines to the left to form the phrase “mumbo jumbo”. Simply stated, the problem is that suffixes are extensions of phrases to the right, so it is hard to look to the left. This problem was solved, however, by Abouelhoda et al. (2004), who added a Burrows–Wheeler transform table to their extended suffix array data structure, giving this this data structure properties of suffix trees. One weakness of Abouelhoda et al.'s approach, however, is that it

does not adapt well to large alphabets. This is, of course, a serious weakness for use in text processing, where one wants at least to work with some subset of Unicode, or even worse, to treat each tokenized word as an alphabet symbol. Fortunately, the restriction to small alphabet size has recently been eliminated in the approach of Kim et al. (2008), who deal with the large alphabet by using binary trees, which are linearly encoded using the child table of Abouelhoda et al. along with a longest common prefix table (lcp).

Using extended suffix arrays makes it possible to count different kinds of occurrences of phrases in different ways. To begin with, we are only interested in counting phrases that repeat. In the text $S = \text{"to be or not to be"}$, the occurrence of the phrase "to be" at $[1, 2]$ is said to be a repeat since the same sequence of tokens occurs at $S[5,6]$ [2]. The difference is that Abouelhoda et al. apply the terms to a pair of occurrences, whereas we apply the terms to a single occurrence. An occurrence of a phrase $S[i,j]$ is left maximal if the longer phrase $S[i-1,j]$ is not a repeat. Thus, for example, the phrase "to" at $S[1,1]$ is left maximal since the phrase at $S[0,1]$ is not a repeat [3]. Similarly, an occurrence of a phrase at $S[i,j]$ is right maximal if $S[i,j+1]$ is not a repeat. If an occurrence of a phrase is both left and right maximal, then the occurrence is said to be maximal. Note that the occurrence of the phrase "or not" at $S[3,4]$ is maximal, though it is not a repeat. Since non-repeats are rarely of interest, we generally assume that we are talking about repeats unless otherwise stated. A phrase is also said to be maximal in a text if there exists a maximal occurrence of the phrase in the text. For example, in the text "mining engineering", tokenized by characters, the phrase "in" is maximal since there are maximal occurrences at $S[2,3]$ and $S[11,12]$. But the longer phrase "ing" is also maximal since it occurs maximally at $S[4,6]$ and $S[16,18]$. So the occurrence of "in" at $S[16,17]$ is a non-maximal occurrence of a maximal phrase. A maximal repeated phrase that is not a subsequence of a longer maximal repeated phrase is said to be supermaximal. Thus the phrase "ing" is supermaximal in this text.

Generally, we are only interested in counting occurrences of maximal phrases since a phrase that never occurs maximally is unlikely to be of interest. But what kind of occurrences should we count? Should we count all occurrences, or only the left maximal, right maximal or maximal occurrences? The answer is that we don't need to decide ahead of time. We can simply test each of these four cases for distinctness, and choose the most distinct case. Take, for example the word "side", which is a maximal phrase in our texts. Should we count all instances of this phrase? Or should we perhaps restrict the count to right maximal occurrences so as to avoid counting those instances that are extended to the right to create the longer phrase "side effect"? Or maybe left maximal occurrences to avoid the longer phrase "flip side"? Or perhaps we should restrict in both directions to avoid either kind of extension. Since it is not generally possible to predict which is best, the reasonable approach is to try all possibilities to see what works best.

One counterintuitive feature of our approach is that it also makes sense to count 0-grams. A left maximal occurrence of a 0-gram, for example, must have a hapax legomena to its left, and a maximal occurrence of a 0-gram must have hapax legomena on both sides. These sequences of two hapax legomena may well be distinctive, since they often are an indication of a named entity or a foreign phrase. Counting all occurrences of the empty sequence is, of course, equivalent to counting the text length, which may well also be a distinctive feature.

5.2.4.2 Experimental setting for LSA based positioning

The experiments described in this section demonstrate the use of language technologies involved in the implementation of the LSA based approach for positioning. The experiments compare training set results obtained with the traditional bag of words LSA configuration against the alternative configuration that uses maximal phrases as the unit of analysis. The alternative LSA configuration starts with a vector space model that (instead of using words counts) uses counts of distinctive phrases that occur at least once as a maximal phrase within the text collection under analysis.

5.2.4.2.1 Data

Our ongoing work in word co-occurrence models for learner positioning extends the existent LSA based approaches and is aimed at analysing and then scoring texts posted on an online medical student discussion forum, where University of Manchester students discuss issues related to one of 6 subtopics of the general topic of safe prescribing. We built a training set consisting in 504 postings that were annotated by experts with four grades (i.e. 109 poor, 200 fair, 142 good, 50 excellent) and one of six topics (i.e. 42 of topic 'a', 50 of 'b', 130 of 'c', 22 of 'd', 247 of 'e' and 13 of 'f'). Each grade is based on the individual posting's textual contribution to a series of expected learning outcomes. Highly scored postings can then be used as evidence of learner proficiency in the corresponding topic.

5.2.4.2.2 Building the bag of words and phrase based vector spaces

As already explained, to identify and extract the maximal phrases we analyse suffix arrays using an extended version of the Yamamoto and Church algorithm to generate all n-grams from a text and avoiding the combinatorial explosion by grouping these n-grams into equivalence classes. Each phrase was counted in one of 4 ways: all instances, left-maximal, right-maximal and maximal. To avoid an unmanageable level of sparseness we include in the analysis all instances of all phrases that occurs at least one time as maximal. Phrases are sorted by their scores absolute values. We then built a 19730 phrases to 504 chat texts matrix that contains the frequency of occurrence of each phrase in each texts. We then weighted the matrix using the tf-idf weighting scheme. We then generate three LSA semantics spaces by reducing the SVD resulting matrix singular values to 50, 100, and 200 respectively. In addition we created another set of 3 bag of words based LSA vector spaces using the same weighting scheme and respective number singular values. In this case using a 6320 tokens to 504 chat texts matrix. The number of token used is the results of choosing the tokens that occurs at least two times within the chat texts collection.

5.2.4.2.3 K Nearest Neighbour based classification

The k Nearest Neighbours algorithm (kNN; Cover and Hart, 1967) is a learning algorithm that classifies texts on the basis of a measure of distance (e.g. cosine) between them. The algorithm classifies each text by looking at k of its nearest neighbours and then assigning it to the most common category represented by those neighbours. If no class is associated to a majority of neighbours, the text is assigned to the category represented by texts with higher cosine similarity. We arbitrarily used a low k value (i.e. k=5) as we expect that noise from the semantic space will be reduced by means of LSA. A common criticism of kNN is that, since it doesn't make any generalizations from the data, it is prone to overfitting. We assume that this criticism should not apply completely to vector spaces generated by means of LSA as the SVD dimensional reduction smoothes and therefore reduces the effect of over fitting that is usually present in kNN based classification.

5.2.4.2.4 Training set kNN results

Experimental results shown in Figure 6 demonstrate that for some topics and grades using maximal phrases as units of analysis can improve the performance of LSA. The x axis corresponds with the different LSA and k-NN configurations implemented for training purposes e.g. D50 indicates that 50 LSA dimensions were used, w or p indicates if phrases or words were used and the last value corresponds to the number of neighbours used in the kNN algorithm. The y axis indicates the fraction of correctly classified documents for each of the classes (grades or topics). In fact, the best results for two of the four grades (e.g. excellent and poor) were yielded by vector spaces built from phrases. Different results are produced by kNN classification for topics where spaces built from bags of words produced the best results clearly for at least four of the six topics representative of their class.

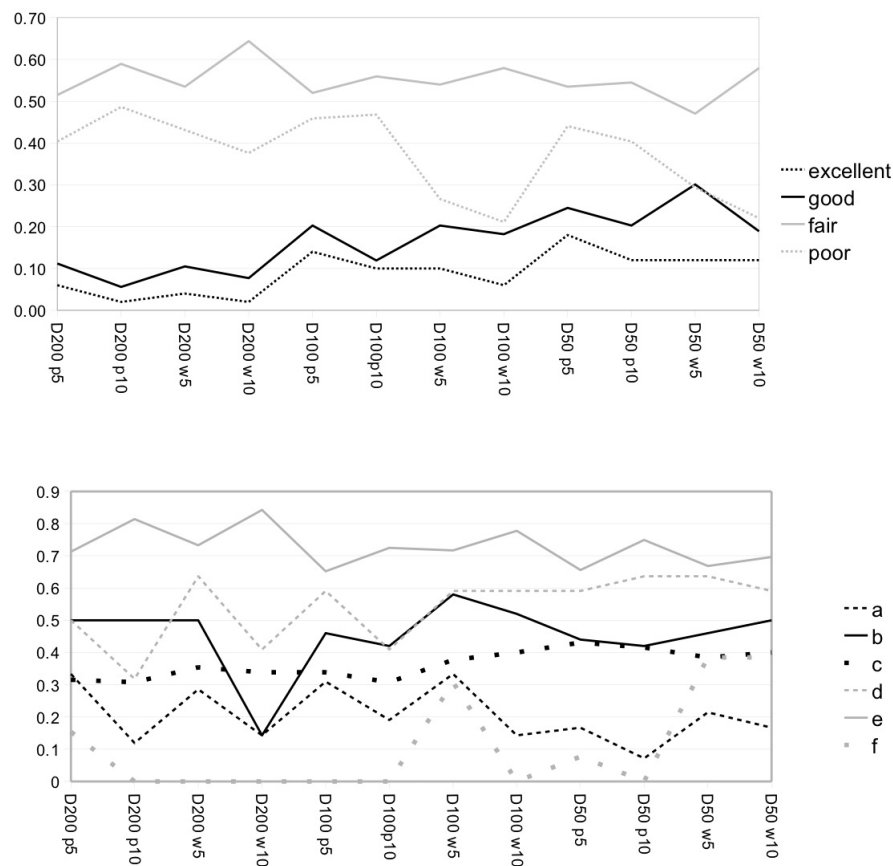


Figure 6: Training set results for kNN topics and grades categorisation

For particular grades and topics, the phrase based LSA (i.e. using vector spaces built from phrases occurring at least one time as maximal) appears to improve over LSA results that have been obtained with the traditional bags of words approach. These results are encouraging and therefore we plan to test alternative semantic space configurations in particular using more distinctive phrases (e.g. all maximal, left maximal and right maximal). We expect that as we collect a larger text sample to build the training set, we will be able to afford the use of those phrases without facing unmanageable levels of sparseness detrimental to results already obtained.

5.2.5 Conclusion

Our results at this point are rather tentative. They seem to show that the bag-of-phrases approach may work better than the bag-of-words approach. But there are still many parameters that can be experimented with. Therefore more definitive results may be forthcoming. In particular, different approaches for measuring the distinctiveness need to be further investigated, both for accurate text classification and for qualitative feedback for the user. Moreover both the quantitative and qualitative feedback need to be carefully validated to confirm that this is an effective service for practical use.

5.3 Knowledge rich approaches for positioning the learner

The learners within some community of practice can position themselves with the help of various knowledge-rich resources, viewed as target competence providers. These resources might be shared among community members and/or recommended by a community member, or even developed by them. For example, it might be a domain ontology, a terminology lexicon, curriculum and learner's profile, annotated with concepts from the ontology. As stated in (Wenger 2001:39): "Communities of practice do produce and share documents and other knowledge artifacts, which can be put in electronic form, and which they need to manage effectively."

Concerning the Stahl cycle, this subtask can be placed predominantly in the space of cognitive artifacts (upper left part of the table), because it uses knowledge rich resources and methods to do the positioning. Thus, on the one hand, it complements knowledge poor subtask, which is more collaboratively oriented. On the other hand, it can be viewed as a transition to WP4.2, because it uses some kind of reference model for comparison. In our case, this is a domain ontology which is adapted to the needs of our particular stakeholders. As Bowker and Star (1999) have shown, ontologies are not absolute but are rather dependent on the world view of particular communities. As such world views change, ontologies need to adapt. One currently popular approach for achieving a continuously adapting ontologies is to use folksonomies, as is being investigated in WP6.2. But since folksonomies are democratically determined by users of many different backgrounds and education levels, they are less appropriate for the positioning task.

In order to avoid terminological misunderstandings among tasks, we will give a definition for the term 'concept' in our positioning task. It is as follows: formalization of a class of objects, specifying relations with other objects and their properties. The concepts are defined in a formal ontology. Additionally, the ontology contains definitions of relations and some instances.

The term general ontology might mean either 1. upper ontology, or 2. linguistic ontology, or 3. how to handle prototype conceptualizations in various environments (groups, cultures, etc.). Thus, in order to achieve completeness, our answers cover these three possibilities:

1. Upper ontology is the upper part of any ontology. The relation is that the domain ontology inherits information from this upper part. The upper part also supports the reasoning and consistency of the domain ontology.
2. Linguistic ontology is any lexicon, organized with respect to some taxonomical structure. These types (e.g. wordnets) are considered as thesauri by (Guarino 2000).

Their relation to the light ontologies, at which we aim, are as follows: they provide the lexicalizations of the concepts, presented in the ontology. Also, they support the connection between the domain ontology and the upper one, i.e. they cover a very important part – the middle one. This part is a bit more abstract than the domain concepts, but less abstract than the upper ones. In this way, it is more convenient to work with, when connecting lexica to ontology, and when extending ontologies.

3. Prototype conceptualizations can vary among various groups of interests, age, nationality, etc. However, we aim at covering the most important concepts in a certain domain. Partly, we handle the interpretation problem with the lexicons, mapped to the ontology. Via a filtering mechanism, the stakeholder can choose whether to use the expert terminology, or the more common lexicalisations.

Knowledge rich approaches are considered in two ways: (1) a separate module for learner positioning; and (2) a module which provides information for the LSA-based methods to use concepts instead of terms for positioning. Knowledge-rich approaches provide a more robust way for explication of learner and curriculum competence, because concepts within ontologies integrate all the term lexicalisations in some domain. Thus, this method is supplementary to LSA, which recognizes mentioning of concepts in texts that are not in the ontology. In our work on the positioning of the learner we rely on the ideas reported in (Kalz et al., 2007). They discuss the notion of learning networks, considered within the task as a community of practice. According to this notion, the learner's competence can be automatically compared to a set of concept evidences of the target competence. Our goal is to achieve an ontology-based positioning where the learner competence is represented by a learner competence ontology and a curriculum competence ontology. However, reliable competence ontologies are still missing. Thus, in our work we will rely on domain ontologies which reflect the knowledge part of the learner's competence. The ontological analyses of the learner's profile and the textual description of the relevant curriculum will be an approximation of learner's competence and curriculum competence. The domain ontology and/or curriculum might be provided and/or recommended by an expert within the community of practice. Learner profiles (CVs, interests, suggestions, opinions, comments) are shared within the community. Thus, the expert(s) (designated as such within the group) might help the others with the annotation of profiles and comparison with the curriculum via an ontology. This help might be of different types: the expert has enough expertise to do it himself, or he can contact the appropriate people outside to do that for the community. The curriculum might also be developed by the community. In order to introduce some first steps in the evaluation of the learner's knowledge degree, we will evaluate the usage of this knowledge represented within the profile. This evaluation will be done via the techniques of the sentiment analysis. After analysing the curriculum description and the profile, we will use several approaches to compare the extracted conceptual information. In the rest of this section, we discuss the envisaged analyses in the process of profile analysis.

Knowledge rich methods rely on analysis of the text by using knowledge sources outside of the text, such as lexicons, ontologies, grammars. For this reason, this task is placed in the cognitive part within the Stahl learning model. In the case of profile analysis, the result from it is used as an evidence of the learner competence and knowledge in the domain. Within the framework of work package 4 we consider several types of text analysis (also their applications to learning tasks, and more specifically the positioning of learner with respect to a curriculum description). These text analyses include: (1) ontology-based semantic

annotation, (2) discourse segmentation, (3) lexical chains approach to disambiguation of concept annotation and (4) sentiment analysis for evaluation of the concept usage in the text. These facilities are envisaged to serve communities of practices in a specific domain of interest. The combination of the above mentioned analyses has to explicate the conceptual content of the profile which is to be used for positioning of the learner. The ontology-based semantic annotation relates the text of a profile to the conceptual information in the ontology. The discourse segmentation facilitates the creation of lexical chains and the sentiment analysis. The lexical chains identification supports the disambiguation of the ambiguous terms and phrases within the text. The sentiment analysis determines the attitude of the learner to the concepts explicated within the profile. At the end, the result of the whole set of analyses presents a classification of the concept usages in the text as known and unknown. This classification will be used for the comparison to a conceptual representation of a curriculum via an ontology. Here are some recent references to relevant works. Galley and McKeown (2003) present the idea of the lexical chains, which trace the cohesion within the texts. The automatic establishment of lexical chains is outlined in Mihalcea (2007). Wolf and Gibson (2006) propose 11 coherence relations, which successfully segment the discourse within the texts. According to Schauer (2000), 15 to 20 percent of coherence relations are signalled by some kind of conjunction but not all of them are unambiguous.

The various levels of sentiment analysis scope are described in Moilanen and Pulman (2007) and Liu (2008), among others. It is often emphasized that adding knowledge rich features improves the results in sentiment analysis. For example – Moilanen and Pulman (2007), Kennedy and Inkpen (2006), Kim and Hovy (2006).

As mentioned above, knowledge rich approaches are usually connected with the availability and the usage of knowledge rich data bases, such as ontologies and lexicons. The ontologies reflect the conceptualizations in some domain of interest. For example, the DAML ontology library, SWOOGLE, or the LT4eL ontology. These ontologies have to be connected to an upper ontology in order to cover in better way the general knowledge. For example, DOLCE, SUMO, SIMPLE. The most famous knowledge rich lexicons are the so-called wordnets (WordNet, EuroWordNet, BalkaNet, SIMPLE). Such resources are exploited for semantic annotation of documents and/or for semantic retrieval. For better semantic annotation and its usage in positioning of the learning task, we consider discourse segmentation and sentiment analysis methods as relevant. Once available, the resources might be changed with respect to community's needs by the community itself or by the providers. The sentiment analysis might be done within the community, using only its members' opinions, or in general – using other people's opinion outside community.

Within the LT4eL project, an ontology-to-text relation was defined (Simov and Osenova 2007; Simov and Osenova, 2008). We briefly present this relation here. We assume that the ontology is the repository of the lexical meaning of the language. Thus, we have started with a concept in the ontology and we searched for lexical items and non-lexical phrases that convey the content of the concept. There are two possible problems here: (1) there is no lexical item for some of the concepts in the ontology, and (2) there are lexical items in the language without a concept representing the meaning of the lexical item in the ontology. The first problem is overcome by allowing in the lexicon also non-lexical (fully compositional) phrases to be represented. The second problem is solved by extension of the ontology. The lexical items are then mapped to grammars. We call them concept annotation grammars. These grammars relate the lexicon to the text. Such a mapping is necessary, as many lexical

items and phrases from the lexicons allow for multiple realizations in the text. Thus, they require some additional linguistic knowledge in order to disambiguate between different meanings of some lexical item or phrase. Figure 7 depicts the elements of the model.

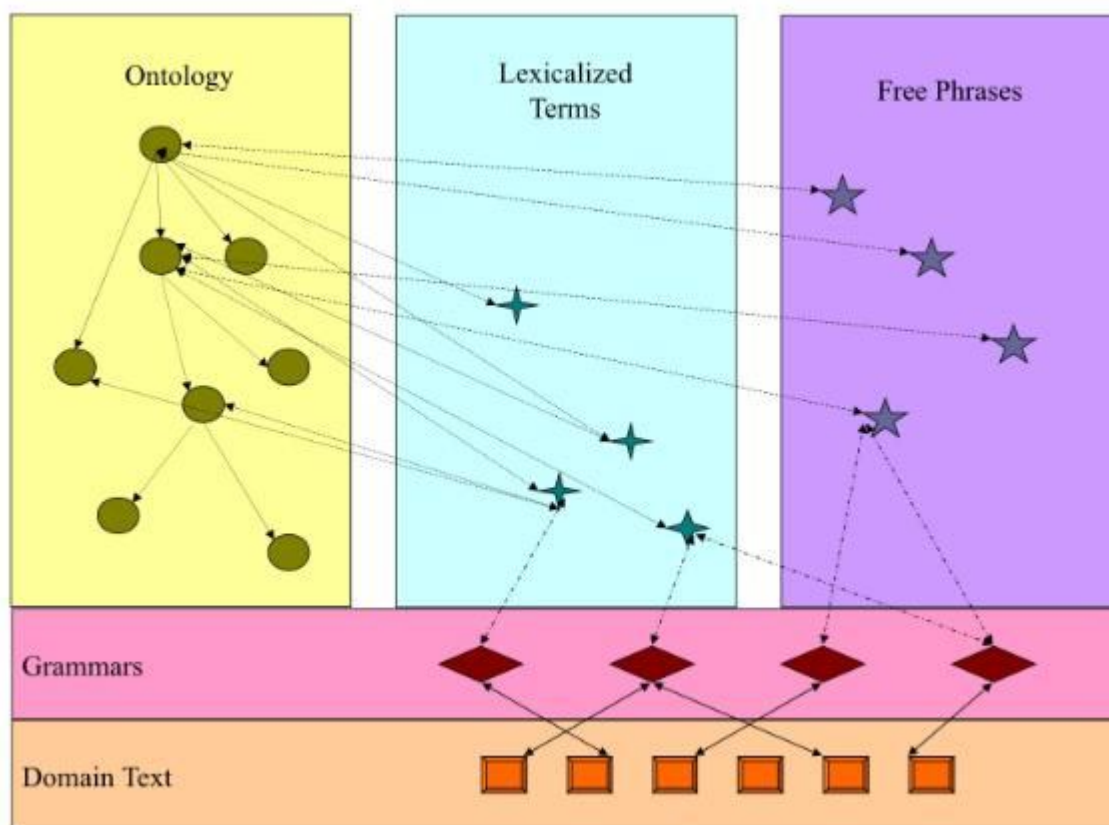


Figure 7: Ontology-To-Text relation to be exploited in the knowledge rich approach to learner positioning

We have been using the relations between the different elements for the task of ontology-based search. The connection from ontology via lexicon to grammars is relied upon for the concept annotation of the text. In this way we have established a connection between the ontology and the texts. The relation between the lexicon and the ontology is used for definition of user queries with respect to the appropriate segments within the documents.

Another direction of the knowledge rich methods is discourse analysis. As a benchmark, the work of Wolf and Gibson (2006) can be considered. The authors present 10 coherence relations. The advantage of their work is that they succeed in collapsing the large number of possible relations into a small set of operational relations.

Sentiment (opinion) analysis can be specified broadly as a kind of analysis that aims to determine automatically the attitude (sentiment, tone, polarity) of a speaker/writer with respect to a certain topic. Usually this kind of analysis is opposed to the standard fact-based analysis and at the same time it is rendered as a classification task. It is commonplace to

evaluate the sentiment of an opinion on a two-value scale: negative or positive. When free of subjectivity, the text is regarded as neutral.

Much work is done for detecting negative or positive judgment, but sentiment analysis is not only about the more general polarity of an opinion, it is also about identifying the opinion holder, the object (the topic) that have been evaluated, the type of propositional attitude expressed (belief – think, believe, assume, emotion – hate, adore, etc.), and the strength of the polarity. Of course, more features can be added.

The type of texts subject to this kind of information retrieval are numerous, just to give some examples: customer reviews on all kinds of products, brands, reviews on cultural events, opinion polls. Here are some freely downloadable corpora with different domain-specific texts: (1) MPQA Opinion Corpus, containing 535 news articles collected during the 2002 NRRC Workshop on Multi-Perspective Question Answering (Wiebe, 2002); (2) polarity dataset, containing 1000 positive and 1000 negative movie reviews (Pang and Lee, 2004).

Sentiment analysis is concerned with two levels of granularity: sentence and document level, the second estimated as too coarse for most applications (Liu, 2008). According to Moilanen and Pulman (2007) though, while sentiment classifiers work well with a large input (e.g. a 750-word movie review), the results for sentential and subsentential units – clauses or noun phrases, are not satisfying. Taking into account linguistic features, such as valence shifters (for example negation) intensifiers, gradable adjectives, patterns, semantic role labelling and syntactic structure, adding a level of compositionality do improve the analysis in terms of accuracy (Moilanen and Pulman, 2007; Kennedy and Inkpen, 2006; Kim and Hovy, 2006, amongst others).

The prevailing majority of techniques use some form of machine learning, supervised and semi-supervised. The most common and basic approach to sentiment classification is keyword-based, starting from a list of sentiment indicators or clues prepared manually, semi-automatically (relying on WordNet or FrameNet) or acquired by machine-learning (Rimon, 2005). Other types of elements used in different algorithms are: Semantic Orientation (SO) – Pointwise Mutual Information (PMI), support vector machines, maximum entropy, naïve Bayes, latent semantic analysis (LSA) and so on (for a detailed overview see Liu, 2008).

In this knowledge rich approach for learner positioning we will rely on the reported works. We will integrate the above technologies into a common processing module in order to explicate the conceptual content of the profile and the curriculum description. The explication of the conceptual content will be done via annotation of the text part of the profile and the curriculum description. This annotation could be used as input for different tasks. Firstly, the concept annotation will be used to find the position of the learner with respect to the curriculum and to select appropriate learning materials to cover the gaps discovered by the method (see below for more details). Secondly, the concept annotation within the text could be used as an input for LSA methods. Concepts could substitute the terms within the vector space.

To evaluate the performance of this knowledge rich approach for positioning we plan to use comparable data sets and validation method as the ones proposed for the LSA based approach.

5.3.1 Extending knowledge rich approach tools

As mentioned above, additionally to LSA based functionalities, the service will provide knowledge rich profile positioning functionalities. For that purpose we plan to extend the CLaRK system (Simov *et al.*, 2001) originally implemented with the aim of minimizing human work during the process of corpora creation. CLaRK will provide service's functionalities for calling external programs when they are necessary for some specific task. CLaRK is implemented in Java and the necessary functional interface will be provided.

The knowledge rich methods which are envisaged to be implemented by the profile analysis and positioning service are as follows: (1) semantic annotation of the profile, (2) lexical chains approach to disambiguation of concept annotation (3) discourse annotation of these texts; and (4) sentiment analysis of the discourse segments as well as the mentioned concepts with respect to the levels of learner's concept competence. They will be combined in a common procedure. The result of the knowledge rich analysis of an profile will be a concept evidence of the learner's competence expressed in the profile. The elements of the concept evidence of the learner's competence will be a set of concept descriptions extracted from the profile with links back to the text of the profile. In this way the concept evidence of the learner's competence can be automatically compared to a set of concept evidences of the target competence (learning network in the terms of Kalz *et al.* 2007). Those will be selected that are not covered by the current learner's competence. For the comparison of the concept evidences we will use the standard vector metrics from Information Retrieval community. The links to the profile will support the assessors of the student competence to find out the reason for the inclusion of a concept description in the concept evidence of the learner's competence. The content analysis which is meant to be implemented for this task will allow us to use the methodology for positioning of learners presented in Kalz *et al.* (2007). Concept descriptions used for the semantic annotation and for the representation of concept evidences are taken from the domain ontology. Recall that here we consider only an approximation of the learner's competence based on the concepts from a domain ontology and their usage in the profile. Much more work will be necessary in order to support a full representation of the learner's competence. The same applies to the target competence encoded in the curriculum description.

The semantic annotation and the discourse annotation will be used also in work package 6 (WP6). The difference will be in the domain of application and the specific type of text which will be analysed here, namely the profile document.

5.3.1.1 Semantic annotation

In order to use the LT4eL model for the analysis of the profile we will implement the ontology-to-text relation for the new domain (medicine) with a new vocabulary. We will extend the previous implementation with new disambiguation functionality which will be based on lexical chains (Galley and McKeown, 2003), using semantic annotation of general words in the text (in addition to the domain specific terms) and discourse annotation. For the semantic annotation of the general words we will use OntoWordNet (Gangemi *et al.*, 2003), which is already aligned to the same upper ontology which will be used in the construction of the domain ontology. The output of this new functionality will be a semantically annotated text of the profile. Each domain term will be annotated with a concept from the domain ontology and each general word will be annotated with concepts from the upper ontology.

5.3.1.2 Discourse annotation

Similarly to the task within WP6, our main goals in developing an additional layer of discourse segmentation and relations annotation are: (1) to investigate the possibility for refining concept recognition and sense disambiguation for targeted words (lexical terms) via coherence relations (discourse relations, rhetorical relations) markup; (2) the discourse annotation will be used for the sentiment analysis in order to evaluate the learner's attitude to the concepts mentioned within the profile. The input for this functionality will be the results from the previous functionality. The discourse annotation process consists of several steps, which may be iterated.

1. Creation of coherence relations taxonomy. We will start with the set of relations and the coding scheme defined by Wolf and Gibson (2006). Their taxonomy is based on the Hobbs list (Hobbs, 1985), and is more coarse-grained than others that include up to 400 types of relations. This fact makes it really applicable for our task. It consists of eleven types of coherence relations:

Temporal sequence: When one discourse segment describes an event that takes place before another event, expressed in another discourse segment.

Cause-effect: When one discourse segment describes the cause, and another – the effect for a given event.

Condition: When one discourse segment describes a possible event that will occur only if another event, described in another discourse segment, also occurs.

Elaboration: When one discourse segment elaborates, i.e. gives more detailed information about another discourse segment.

Example: When a discourse segment provides examples for another discourse segment.

Similarity: When the event, expressed in one discourse segment, is similar to an event, expressed in another discourse segment.

Contrast: When the event, expressed in one discourse segment, contrasts an event, expressed in another discourse segment.

Generalization: When one discourse segment states a generalization for the content of another discourse segment.

Violated expectation: When there is an absence of a causal relation between two discourse segments.

Attribution: When one discourse segment states the source for the content of another discourse segment. It is usually used in constructions, such as: *John said that...*

Same-segment: Same-segment is a structural type of relation, because it holds between disconnected parts of one discourse segment (subject NP separated from its predicate). Same-segment, similarity and contrast relations are symmetrical while the

rest are asymmetrical (directed), that is – one of the segments is more important (the nucleus) than the other (satellite).

According to the coding scheme the three general steps of the annotation process are: (1) the output of the sentence-splitter is segmented further into clauses and then, if needed, annotators insert intrasentential boundaries for smaller discourse segments; (2) the discourse segments are grouped thematically and (3) the coherence relations between the segments are indicated. After the annotation process is finished, the taxonomy may be further adjusted to improve the descriptive adequacy for the texts in the Computer Science domain.

2. Manual annotation of portfolios as a gold standard for the creation of automatic procedure for discourse segmentation.

3. The analysis of the obtained discourse structures will provide information that could be used for (1) the development of constraints over the semantic annotation grammar, (2) supporting anaphora resolution, and (3) support of sentiment analysis. In addition, we consider the possibility of creating a rule-based grammar for recognizing the coherence relations that are unambiguously linguistically marked.

In order to improve the concept annotation, we will test different knowledge-based techniques that are common for word sense disambiguation. Our main goal is the enrichment of the concept annotation grammar in order to map the relations between text chunks, recognized as carriers of the concepts, with relations present in the domain ontology: is-a, part-of, etc. In the future different algorithms for automatic establishment of lexical chains (with nouns) may be tested (for an overview see Mihalcea, 2007). Lexical chains and rhetorical relations, the two types of discourse information, contributing to the text coherence, will be used for improving the concept annotation. For example, a discourse segment, nucleus in an elaboration relation, will most probably contain a term, connected via hypernymy relations with lexical units that belong to the satellite segment.

The discourse annotation will be adapted to the format of the profile. The idea is that elements of the profile will require some specific kind of language. In such cases the discourse structure might depend on the peculiarities of the corresponding sub-language. The output of this functionality will be a segmentation of the text of the profile in discourse elements and annotation of the relations between them.

5.3.2 Sentiment analysis

The input for this functionality will be the results from the previous above described functionalities. In order to construct a concept evidence of the learner's competence, we first need to extract the concepts which are mentioned within the profile. Then, on the base of the ontological reasoning, the implied concepts will be added. For example, if the profile's holder in IT domain says that he/she has some expertise in XSLT, this automatically means: on more general level, that he/she has also knowledge of XML and some programming language, and, on more specific level, that he/she can use XML-based language for the transformation of XML documents into other XML or "human-readable" documents. We also need to know in what context each of the concepts in the profile was mentioned by the learner. For example, behind the discourse relation, called *contrast* the learner stated two opposite facts: it is useful to know how to transform documents, but a next step is required – to learn also XSL-FO language in order to handle formatting objects. From this short context a conclusion can be

drawn that the learner's position with respect to the knowledge of XSL set of languages is partly completed. Thus, comparing conceptual information and discourse relations, each mentioning of a concept will be evaluated by one of the values: 'well known', 'known', and 'unknown'[4]. We will use the methods developed in the areas of sentiment and opinion analysis. As it was already mentioned, a pre-defined requirement list of necessary concepts with definitions will be used in order to estimate the degree of competence, delivered by the learner in the profile. There will be three types of evaluation: coverage, degree of detail and relevance. The coverage will be estimated over the number of the mentioned relevant concepts that match the pre-defined list. The degree of detail will be evaluated over the depth of the conceptual space. And the relevance will be estimated via the ontological relations from a given concept to the other co-occurring concepts within the discourse segment.

5.3.3 Construction of a concept evidence of the learner's competence and knowledge

As it was mentioned above, a concept evidence of the learner's competence is a set of concept descriptions extracted from the profile. For the moment we divide this set in the following subsets: (1) known concepts; (2) partially known concepts; (3) unknown concepts; and (4) concepts with contradictory usages. The first subset will contain all the concepts which are evaluated as known in the profile. The second subset will contain concepts that are mentioned in the profile, but for which there is not enough evidence about the level of knowledge of the learner with respect to them. The third subset will contain concepts that definitely are mentioned as unknown by the learner. In the last subset we will include the concepts for which there are positive and negative evidences about the knowledge of the learner. In addition to the extracted concepts we will extract links to the occurrences of the concepts in the text. Within the community of practice, the curriculum part has to be defined against which the positioning to be done. For example, the curriculum might take XSL as a whole set of languages, in which each language (XSLT, XPath and XSL-FO) has to be learned. On the other hand, only XPath might be taken as a learning goal.

The output of this functionality will be used further to compare the concept evidence of the learner's competence with the community of practice. The comparisons will use a vector representation of concept evidence of the learner's competence and concept evidence of the target competence. The vector for target competence will be fixed within the learner network. The vector for learner's competence will be created by the assessor on the basis of the above sets of concepts.

The evaluation of the method will be done on two levels. First, for each of the processing steps, we will create manually gold standard corpus on which to test the corresponding technology using the usual precision and recall metrics. Second, we will test the method with respect to the performance of the LSA-based method. The aim is not for the methods to compete, but to find the best ways to combine them in order to satisfy task goals.

We have described the knowledge rich method preferably as a complement to LSA, rather than an alternative. We envisage also integration of the two methods. First, in the construction of a vector space instead of terms from the text the concepts from the conceptual annotation could be used. In this way, one can abstract over the textual representation of the concepts. For example, very often in text a super-concept term can be used to denote a sub-concept – "system" instead of "computer system". Also with the sentiment analysis we could select which concepts to be included in the vector space. It is also possible to combine the two methods via integrating their results.

6. Diagnosing conceptual development (task 4.2)

6.1 Outputs of WP4.2

WP4.2's contribution to the integrated environment will be an elaborated scenario series developed in association with WP3, and services corresponding to the scenario. These services are described in more detail below and briefly comprise services for diagnosing conceptual development and functionality for aggregating the raw data output into formats meaningful to the end user.

6.1.1 The WP4.2 solution scenario, providing formative feedback

We have elaborated a solution scenario (see deliverable D3.2) that depicts the functional design of a service aiming to provide (semi-)automated formative feedback with the help of Language Technologies. Using the service, learners could compare evidence of their knowledge (e.g. text inputs such as essays, blogs, "think alouds" etc.) with reference models in order to identify possible differences and obtain recommendations of suggested actions to address the differences. Learners can submit new evidence of their knowledge and receive formative feedback as often as they want. Moreover, learners can monitor their own learning process as the service provides also comparisons of the learner's knowledge evidences previously submitted.

For tutors, the service will provide a means to monitor the current progress of learners on a topic, to allow them to take proactive actions to improve learners' conceptualization of the topic. This might lower tutor workload. The design considers that the service can be used in both formal and informal learning settings. Depending how the use of the service is implemented in the learning context, learners can assume both tutor and learner roles (for more detail, see WP4.2 Solution Scenario included in D.3.2).

The scenario will be implemented as a set of web services providing (1) learner evidence collection facilities (data gathering), (2) data extraction and condensation / aggregation functionality and (3) facilities to compare concepts with the reference models and present the results to the user.

6.2 Research problems

In undertaking this work, the following research problems arise:

- do potential end users find the scenario realistic?
- can we adapt existing concept mapping tools to meet the requirements of the scenario?
- with what reference models should the learner's conceptual development be compared?
- how should the raw data from the service be aggregated to present meaningful information to the learner and tutor, to inform their future actions?
- what tuning of the language technology services is required to optimise the delivery of meaningful information?

In the first phase of our work, we undertook a conceptual validation of the showcase scenario (reported in deliverable D7.2), which showed that learners and tutors do find the scenario realistic, subject to a number of enhancements and clarifications. It was identified that "the

ability for students and tutors to compare concept maps is an under-developed area in the showcase scenario that was perceived as having the potential to add substantial value to the service. This feature of the service needs to be developed further to provide a clear picture of the comparisons that students and tutors will be able to make between concept maps".

This deliverable reports our work with respect to the second and third problems. To address the second problem, we have compared a number of concept mapping tools. The work is reported in Section 6.4

With regard to the third problem, we investigate three types of reference model, against which to compare learner data. The work is reported in Section 6.5.

In later phases, we will address the underlying language technology questions, in cycles of software development and validation with end users. In view of the importance of the user interface being meaningful, more frequent validations with end users (learners, tutors) are indicated.

6.3 Introduction to studies

In our discussion of the theoretical basis of learning, we have indicated the importance of the inter relationship between the individual learner and communities of practice and the interaction between building "personal knowing" and "collaborative knowing". We have also indicated that in order for learners to develop expertise in their specific domain, it is essential that they recognise the limitations of their understanding and conceptual development and develop appropriate learning plans, as demonstrated in reflective learning cycles. The theoretical basis for understanding development of expertise in professional domains identifies knowledge creation and restructuring as essential components of this process. The Stahl model and the notions of Communities of Practice indicate that learning from peers has a key role in enabling individual learners to reach a shared understanding of specific aspects of their domain. The use of Problem Based Learning, which we indicate is used in medical education to model aspects of Communities of Practice, is an example of an educational approach in which peers reach a consensus view of a specific issue, topic or concept.

In order to provide individual learners with the guidance and "instruction" to enable their development of expertise, we require reference models. Within this educational context, we have defined three types, against which learners can compare their understanding of a specific topic. These are:

- Archetypical reference model: based on expert and state-of the art information (e.g. scientific literature).
- Pre-defined reference model (or 'Theoretical reference model'): considers specific information based on the curriculum (e.g. course material, tutor notes, relevant reading materials, etc.).
- Emerging group model: considers the concepts and the relations between those concepts that a group of people (e.g. peers, participants, co-workers, etc.) used most often.

We have concentrated on the pre-defined and emerging approaches to identify or approximate the conceptual development of learners to underpin the role of Language Technology tools. Next, we explain how existing applications and tools, namely Leximancer

(Smith & Humphreys, 2006) and Pathfinder (Schvaneveldt, 1990), have been used in two different preliminary explorations as proof of concept of the suitability of these approaches.

In order to assess the individual's knowledge of a particular domain Goldsmith et al. (1991) proposed a structural approach to determine how the individual organizes the concepts of such a domain. This approach involves three steps: knowledge elicitation, knowledge representation, and evaluation of the representation.

Knowledge elicitation techniques measure the learner's understanding of the relationships among a set of concepts (Jonassen et al., 1993a). Methods that support this activity include card sorting, concept maps, think aloud, or essay questions.

Knowledge representation reflects the underlying organization of the elicited knowledge (Goldsmith et al., 1991). Advanced statistical methods (e.g. cluster analysis, tree constructions, dimensional representations, pathfinder nets) are used to identify the structural framework underlying the set of domain concepts.

Evaluation of the representation relative to some standard (e.g. expert's organization of the concepts in the domain) uses one of the following approaches (Goldsmith et al., 1991): qualitative assessment of derived representations; quantifying the similarities between a student representation and a derived structure of the content of the domain; or comparing the cognitive structures of experts and novices.

6.4 Comparison of existing concept mapping tools - which tools to use?

A decision was made to start the exploration with the cognitive map method, which is one of the most common methods for representing cognitive structures, as a mean to elicit and represent learner knowledge. The decision was taken on the basis of the appropriateness of concept maps for representing learners' representations of subject matter structure and on research evidence that demonstrates the concept map method is well suited for eliciting knowledge (Nesbit and Adesope, 2006), and is a better method for evaluating meaningful learning of learners of different ages than classical assessment methods such as tests and essays (Jonassen *et al.*, 1997; Novak, 1998). It is important to point out, however, that the creation of concept maps is a complex and time consuming task that requires training and practice to understand how the relevant concepts should be identified and how to make the relation between them.

There are already a number of tools for automatic construction and support of the construction of concept maps: Knowledge Network and Orientation (KNOT, PFNET) (Clariana *et al.*, 2006); Surface, Matching and Deep Structure (SMD) (Ifenthaler and Seel, 2005); Model Inspection Trace of Concepts and Relations (MITOCAR) (Pirnay-Dummer, 2006); Dynamic Evaluation of Enhanced Problem Solving (DEEP) (Spector and Koszalka, 2004); jMap (Jeong, 2008), and ProDaX (Oberholzer *et al.*, 2008). Table 2 depicts these tools in terms of the data collection they use, the analysis they perform, the data conversion they use and the comparison they perform.

Tool	Data Collection	Analysis	Data Conversion	Comparison(s)
KNOT	Concept pairs/Propositions	Quantitative Analysis	Pathfinder Networks	Direct comparison of networks with some statistical results.
SMD	Concept map or natural language	Quantitative—analysis is calculated using tools.	Structural decomposition into 3 categories (manual and semi-automatic)	Unlimited comparison
MITOCAR	Natural language	Quantitative—analysis included multiple calculations using tools	Structural composition into 1 category (automatic)	Paired comparisons for semantic and structural model distance measure
DEEP	Annotated causal maps	Quantitative/qualitative—analysis is done mostly by hand	Structure decomposition into 3 categories (automatic)	Unlimited comparisons, showing details relative to concepts
jMap	Concept maps, causal maps, or belief networks	Quantitative analysis – analysis is calculated using tools	Structural decomposition into link strengths between causal factors and evidentiary strength	Superimposes maps of individual (n=1) and group of learners (n = 2+) over a specified target map
ProDaX	Association Data, Cross-Tables, Two-Way Two-Mode Data, Coordinates, Scales	Non-Metric Multidimensional Scaling/Cluster-Analysis	Concept Maps	Comparison of maps based on Procrustean Transformation/Loss-oriented Meta Map (LOMM)

Table 2: Overview of concept mapping tools (adapted from Shute et al., submitted)

These tools have some common characteristics: (a) they are concerned with conceptual development of learners; (b) they can (semi-)automatically construct concept maps from a

text; (c) they use a sort of distance matrices; (e) they propose a quantitative analysis of the maps; and (d) most of them pretend to support high levels of learning, namely critical thinking and problem solving.

Amongst their differences we have found that, even though they all use some sort of Language Technology analysis, not all of them refer to it explicitly. The SMD and jMap can use as an input not only text but also concept maps. These tools also differ in the scoring schemas they use to perform the quantitative analysis: DEEP uses a number of nodes and links; SMD uses propositions or a number of the links of the shortest path between the most distant nodes.

Most of the referred concept mapping tools provide opportunities to identify the conceptual gap between a learner's concept map and a criterion map (in fact, an expert map), or to compare a learner's concept maps in different periods of time. However, only SMD, jMap and, in some extent DEEP, provide purposely a visualisation of this progression towards the criterion. Most of these mapping approaches construct and analyse individual maps. jMap visualises and assesses changes observed in either individual or collective maps. Nevertheless, jMap is the only tool restricted to producing a particular type of maps, causal maps.

KNOT, SMD and MITOCAR do report on reliability and validity criteria correlating, as a typical case, the automatic scores generated by these concept mapping approaches and human concept mapping scores and human essay scores. Finally, it is worth mentioning that SMD and MITOCAR report experimental data on the effectiveness of a particular technique as an increase in similarity between a learner's map and an expert's map.

6.4.1 Methods

A first exploration of existing tools that create concept maps from an input test was performed. The aim was to investigate in which and to which extent existing tools support the process from knowledge elicitation to evaluation. In particular, the aim was to gain insight in how flexible and easy to use the tools are (other aspects such as reliability and validity having been derived from literature).

The following tools have been explored:

- CMAP (Institute for Human and Machine Cognition, 2004)
- KNOT and Ala-reader (Clariana *et al.*, 2006)
- INFOMAP (Peters, 2005)

KNOT is a software tool that generates text proposition files that can be imported from CMAP to generate concept maps automatically. The conclusion from the initial exploration was that the process could be used to analyze conceptual development but there are restrictions on the data that can be used, for instance, the general limit of concept pairs that can be used. Next, an initial exploration of INFOMAP was performed, to generate an associative semantic network, based on learner texts. INFOMAP employs a similar approach to Latent Semantic Analysis, with a focus on word-to-word relations and a context limitation around the words used for indexing.

To this end, a data set from a Psychology course at the OUNL was used. This data offers course content, which was considered as the expert level of argumentation. Also documents

written by the learners were used. For every part of the course content some related keywords were generated, then the inter-correlation of the keywords were calculated for one exemplary chapter. After that, by using a clustering method (nearest neighbour approach), a distance matrix and clusters in the keywords were generated. Figure 8 shows an exemplary cluster overview of a chapter. These keyword clusters can be used to identify topical foci of the documents. An alternative approach to use the keywords and associated other concepts in documents is Multidimensional Scaling (MDS). With this approach distances between concepts can be visualized (see Figure 9).

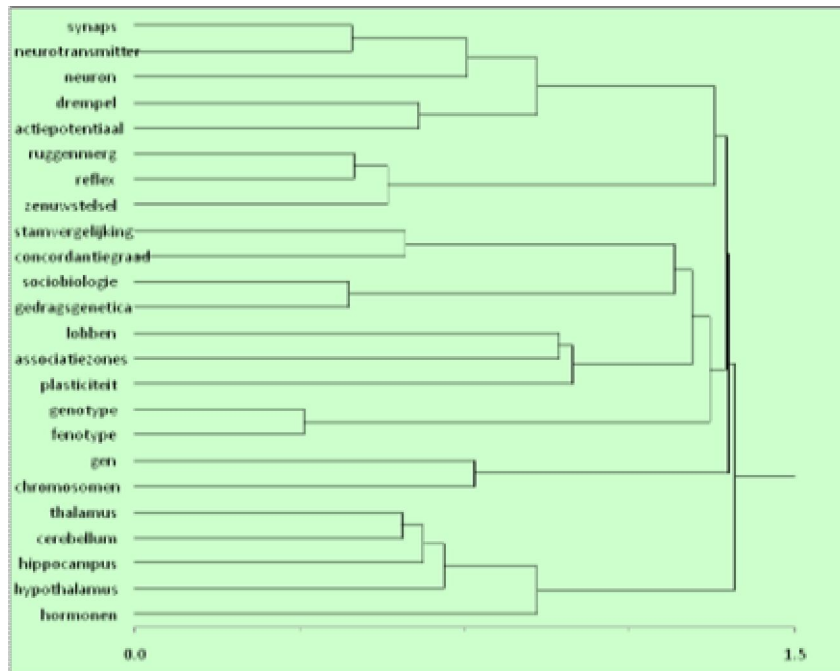


Figure 8: Example of clustering

Derived Stimulus Configuration

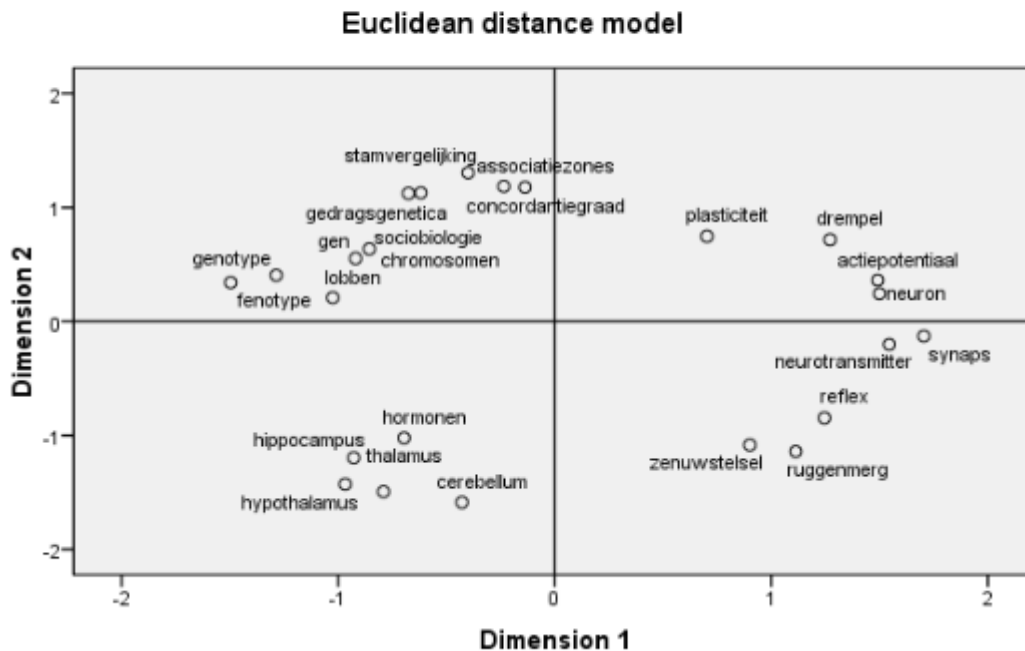


Figure 9: Example of a Euclidian Distance Model derived from Multidimensional Scaling

6.4.2 Results

Clearly Infomap was able to generate a visual representation of the relationship between concepts from the Psychology course texts at OUNL. The use of Multidimensional scaling, however, restricted the use of words associated these concepts and it was therefore decided to investigate tools which non metric Multi dimensional scaling, which allowed the incorporation of a wider variety of language.

Leximancer and Pathfinder were selected for a further proof of concept. Leximancer generates concept maps from a document collection using content analysis (based on co-occurrence) and relational analysis (proximity and concept mapping). These maps, or visual representations, show the concepts identified in the text and the relations between them. Pathfinder can be used to derive and visualize structured (semantic) networks. It is based on proximity measures (similarity, correlations, distances, probability) between pairs of concepts (Clariana et al., 2006).

6.5 Studies on potential reference models using Leximancer and Pathfinder

6.5.1 Pre-defined Reference Model

6.5.1.1 Methods

A data collection protocol was defined to elicit and represent a learner's knowledge. This protocol combines a think aloud procedure with a cognitive map method to provide a suitable and appropriate measure of the learner's representation of the subject matter structure. As a proof of concept these tools have been explored to in two different ways to generate (a) the pre-defined reference model and (b) an emerging group model (see Berlanga et al., 2009 for details).

For the generation of a pre-defined reference model, a combination of Leximancer and Pathfinder was used. A small randomly selected group of Year 2 undergraduate medical students (N = 12) recorded their summaries of a specific PBL case and were asked to speak for 5 minutes on the Bioscience mechanisms which formed the basis of the condition and how they might treat the conditions, using this knowledge. The recording were made under standard conditions in the University of Manchester Medical School, transcribed and used as the text for Leximancer (Figure 10). Text from tutor notes and supporting materials were used to generate the predefined reference model. Pathfinder was used to identify similarities and differences between results from learners and those from the pre-defined model.

The cytotoxic P cells are responsible for killing the microorganisms and it's triggered by the binding of TCR to the MAC protein complex, bound to the specific antigen, the antigen peptide fragments, the T helper cells or the CD 4 T cells are essential for the cell-mediated response. They make cytokines for delayed hypersensitivity and help making B cells specific for antigens. T-regulator cells play a role in the negative regulation of the immune system.

Figure 10: Part of transcribed student think aloud

6.5.1.2 Results

The concept maps from the students and the pre-defined reference model differ in the level of detail (see Figure 11). Whereas the student concept map included detailed concepts, the pre-defined reference model encapsulated the concepts and gave the panoramic view of the knowledge. Furthermore, the student map can be characterized as the description of a disease process, while the pre-defined reference model is at the (auto)immune system level. Finally, the latter includes both a diagnostic part, and more signs and symptoms.

6.5.1.3 Conclusions

These results suggest that even if the learning material explains the reasons and conditions of a problem ("the why"), novice students represent their understanding by indicating only procedural knowledge, mentioning how to solve a problem ("the how"). This might imply that the tutor notes and learning materials might not be ideal to generate a pre-defined reference model. The materials are written from a perspective that requires more expertise than the novice student can achieve at that point of time. Consequently, this might not be a

good basis for deriving the pre-defined reference model, nor for providing formative feedback.

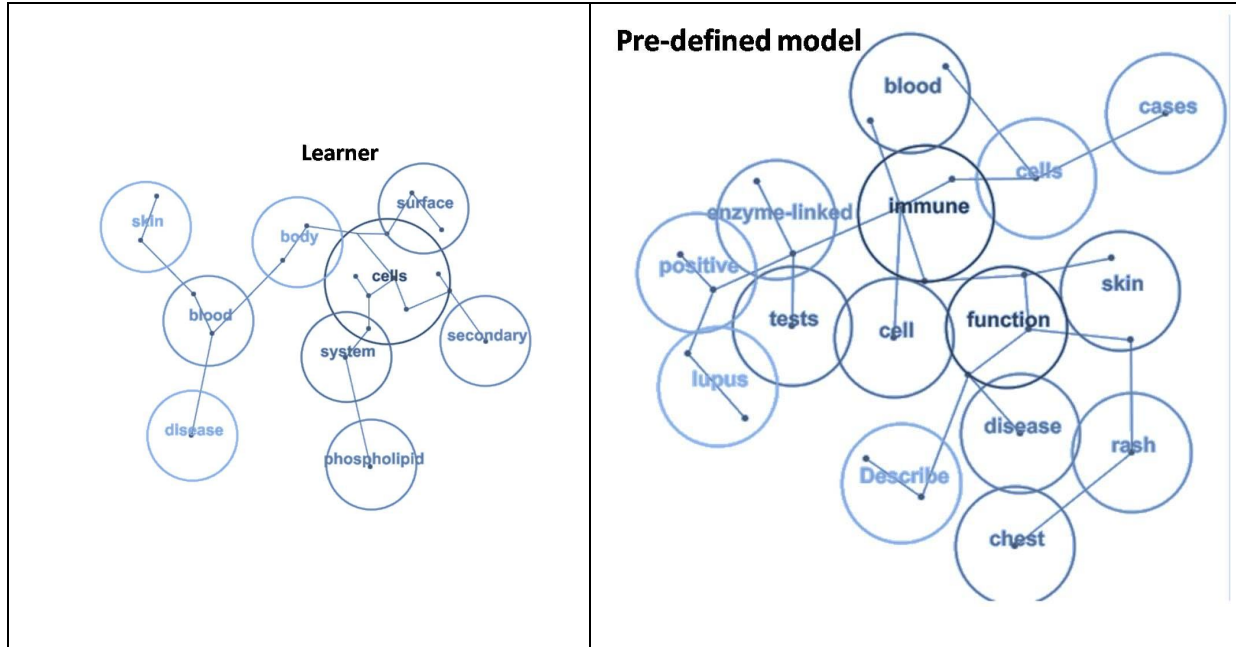


Figure 11: Concept map for a learner (left) and the pre-defined reference model (right) (Leximancer)

6.5.2 Emerging Group Model

6.5.2.1. Methods

Only Leximancer was used in these experiments, in which OUNL employees in an informal learning situation were the subjects. They used a similar “think aloud” protocol to that used in the previous experiments, except that on that occasion the subject matter was “Learning Networks”

6.5.2.2 Results

The results indicated the ten most used concepts and their relevance automatically, as well as the relations of each concept with other concepts. Figure 12 depicts the so-called emerging reference model for the concept *Learning Networks* as it arises from all concepts and the relations between concepts. It also visualizes the position of the individual learners in relation to the model, by indicating which concepts the speaker mentioned.

Further, a concept map was generated for individual employees for whom the ten most used concepts were identified. These were compared to identify similarities and differences between the emerging reference model and employees’ concept maps. It seems feasible to generate individual formative feedback reports that present differences and similarities. Future work involves validation of the reliability of the emerging reference model and the formative feedback report.

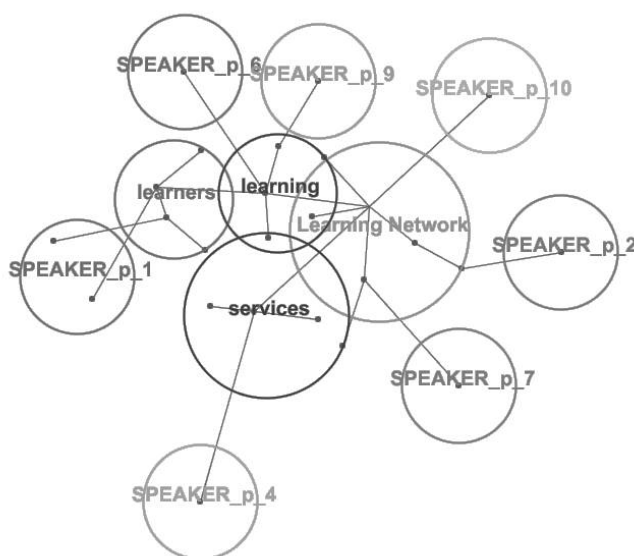


Figure 12: Example of an emerging reference model (Leximancer)

6.6 Conclusions and next steps

In summary, the pre-defined reference model approach seems to provide little information to generate a formative feedback report, since it contained information that might be at a “too high level” for a learner at a specific point of time. In fact this is in line with what we have argued before regarding theories of expertise. It could be the case that, at a specific point of time, learners do not have the expertise level described in the pre-defined reference model, which will consider the ultimate learning goal but not the different levels of expertise a learner will go through. The emerging reference model approach seems to solve this issue. The set of concepts that is used by most people at a specific point in time might provide better evidence of the level of abstraction and relations between concepts. This approach might provide better guidance as it resembles the Zone of Proximal Development (Vygotsky, 1978) of the learner and it could be also seen as a way to build socially a shared understanding of a concept, a unit of understanding that is shared by a particular group and context. The approach, however, will require a better appreciation of the learner’s knowledge representation – by contextualizing both the learner’s knowledge and the situation in which the knowledge will be applied – and requires mechanisms to keep the model updated.

This work informed the development of the WP4.2 scenario. The scenario now makes provision for two reference models (pre-defined model and emerging group model). Undoubtedly, more research is needed to establish how learners would benefit the most from comparing their conceptual development with these models: whether it is good strategy for learners to see comparisons with both models or, whether, depending on their level of expertise, comparisons with different models will be made available. The type of reference model used may depend on the level of learner development. The emerging reference model, which is based on concepts and their inter relationships, generated by peers, would most

likely be of use for an individual learner at a novice level, as at this stage it would correspond to his/her Zone of Proximal Development (Vygotsky, 1978). As expertise develops, the emerging reference model may still be appropriate, depending on the development stage of the practice group as a whole, but pre-defined reference model may be more suited to a more advanced learner.

6.7 Design of the WP4.2 service

While Leximancer and Pathfinder were used in the showcase, they were not considered suitable as a basis for the LTfLL conceptual development service (see deliverable D7.2). Leximancer does provide all the required functionality; however it is a proprietary application and cannot be customized to the requirements of the scenario. Pathfinder also does not provide the required functionality and difficulties were experienced in using it alongside other applications. A decision was made to develop a custom tool based on LSA.

Regarding future work on developing the WP4.2 service, there are three process steps that are needed: (i) data gathering and evidence collection; (ii) extraction of structure and condensation; and (iii) comparison of the conceptual structures (Figure 13).

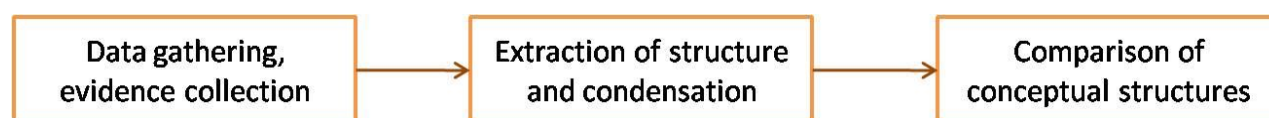


Figure 13: Process steps for monitoring conceptual development

First, **data gathering** and capturing techniques need to provide functionality for easy collection of evidence contained in learning texts using the various tools provided by and beyond the project. Such texts include texts created by the learner, such as study notes, summaries, reviews, discussion articles etc. Further possible evidence sources could include short texts provided in chats, fora, comments, etc. A simple evidence production tool such as a learning diary (blog) as outlined in the section "Texts available for analysis" is also an important source of material from individual learners

The second challenge is the development of web services for the **extraction of structure** into a condensed, meaningful representation reflecting the conceptual information in the learner evidence. Web services will extract the input texts into the desired representational structure, thereby unveiling its conceptual structure. The representation format chosen so far are graphs of connected terms produced by LSA (similar to the concept maps produced by Leximancer, which was shown in the initial studies).

The processing chain is provided as a set of modular services that can be flexibly configured in order to sanitise, tokenise, relate, and aggregate the data from raw input to the conceptual representation at the output. The analytical part of the service condenses the raw information such that the learner gets an overview first and details on demand.

The way in which the data is aggregated is important in determining a meaningful output and the design of the data aggregation will require careful thought. The service will output, for example, selected terms and relations extracted from the text, similarities and differences between the individual terms and the pre-defined reference model, etc.

In the third step, the service will provide a **comparison of the conceptual structures** between the evidence of one learner at a certain point in time and that of another learner, the same person at a different point in time, a pre-defined reference model or a emergent model. This third step faces the challenge of translating the structural representation of the second step into a surface representation of the differences that can be shown and can be understood by the user. As well as considering visualisation methods, navigation and interaction issues will be addressed. The service should therefore help learners to understand the comparison, for example by using visual clues and contextualized help.

For the conceptual development service to be adopted, it is essential that its user interface and reports meet the requirements of the users and provide neither too much nor too little information. Iterative validation of the user interface with potential real users will take place to inform improvements to the outputs. The accuracy of the service is important, i.e. how good the extraction service is compared to human extraction (with dual or more codings to balance inter-rater bias). Technology acceptance testing goes beyond accuracy and evaluates whether the service is appreciated by learners, tutors, and other stakeholders (perceived usefulness and ease of use).

Additional technical information on the showcase implementation can be found in deliverable D2.2. Currently, development progresses towards the version 1 release candidate. Its technology will be documented in D2.3, the stakeholder validation in D7.3, and verification aspects in D4.2.

7. Conclusions

7.1 WP4.1

In summary, WP4.1 supports the positioning of an individual learner by means of qualitative and qualitative feedback that is based on knowledge about learner language usage inferred from the learners' texts and learning materials.

WP4.1 approaches (i.e. knowledge poor LSA based and knowledge rich) rely on identifying evidence of language usage in texts specifically surface level phrases and language usage patterns. While the knowledge poor subtask infers linguistic knowledge from texts only, the knowledge rich subtask makes use of concepts that are determined externally from text by use of ontologies.

This report has described how positioning relates to communities of practice and speech genres and how languages technologies can be used for positioning. The WP4.1 service measures integration in such communities of practise by calculating distances based on textual features and terminology. Moreover, the Stahl cycle has been used to show how integration into expert communities of practice corresponds with level of expertise. For WP4.1, formative feedback comes in the form of commentary on usage of language (e.g. phrases and terminology) that can be used to facilitate effective and responsive communication with experts in the community of practice.

In the next cycle of the project, WP4.1 will focus of on the development of version 1 of the services for positioning based on the design presented here. For supporting the design and the validation of the relevant language technologies for positioning, existing data sets (e.g. graded and annotated online discussion in the medical domain) and new data sets (e.g. graded and annotated computer science text materials in German) are being built. As the data sets are being consolidated, multiple different configurations of the language technologies are being tested. Experiments will be carried out to determine the best use of linguistic patterns (significance of phrases, subject-verb-object grammatical patterns, etc.). Moreover, we expect to experiment with new suffix array algorithms for extracting discontinuous phrases (tandem repeats) as a new source of evidence. Then we will seek to find the best balance between the knowledge poor and knowledge approaches, in order to build a positioning service that can provide the most useful quantitative and qualitative feedback. This system will be validated from the perspective of the user.

7.2 WP4.2

We conclude from our initial studies that we can visualise a learner's ability to relate concepts to one another within a specific domain and compare this to reference models. The type of reference models used, based on either materials from the appropriate curriculum or generated by communities of peer learners, situates these results within Stahl's learning model. Our ability to produce such a comparison in a way that is meaningful to end users will

form the basis for the next stage in production of the services based on WP 4.2, namely provision of feedback to learners, by tutors and other educational stakeholders.

We also conclude that the tools used to produce these results, were not best suited to meeting the stakeholder requirements captured in the scenario. The next step will be the development of a customised approach based on LSA, for which the focus will be both the extraction and aggregation of data. The final format of the service will depend on iterative validation by all user groups.

References

- Abouelhoda, M. I., Kurtz, S. & Ohlebusch E. (2004) Replacing suffix trees with enhanced suffix arrays. *J. of Discrete Algorithms*, 2(1):53–86.
- Alexander, P.A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher*, 32, 10-14.
- Bakhtin, M. M. (1981) *The Dialogic Imagination: Four Essays*, University of Texas Press
- Bakhtin, M. M. (1986) *Speech Genres and Other Late Essays*. Trans. by Vern W. McGee. Austin, Tx: University of Texas Press.
- Bekkerman, R. and Allan, J. (2004) Using bigrams in text categorization. Technical Report IR-408, Center of Intelligent Information Retrieval, UMass Amherst.
- Berlanga, A. J., Brouns, F., Van Rosmalen, P., Rajagopal, K., Kalz, M., & Stoyanov, S. (2009). Making Use of Language Technologies to Provide Formative Feedback. In S. D. Craig & D. Dicheva (Eds.), *AIED 2009 Workshops Proceedings Volume 10. Natural Language Processing in Support of Learning: Metrics, Feedback and Connectivity. Proceedings of the 14th International Conference in Artificial Intelligence in Education, Workshops Proceedings* (pp. 1-8). July, 6-7, 2009, Brighton, United Kingdom.
- Berlanga, A. J., Kalz, M., Stoyanov, S., Van Rosmalen, P., Smithies, A., & Braidman, I. (2009b). Using Language Technologies to Diagnose Learner's Conceptual Development. In I. Aedo., N. Chen, Kinshuk, D. Sampson & L. Zaitseva (Eds.), *Proceedings of the 9th IEEE International Conference on Advanced Learning Technologies (ICALT2009)* (pp. 669-673). July, 14-18, 2009, Riga, Latvia: IEEE.
- Bhattacharya, M. Heinrich, E., and Rayudu, R. (2006). Work in progress: E-portfolios in computer science and engineering education. *Frontiers in Education*, 36th Annual Conference, 11-12.
- Bleakley A, Bligh J (2008) Students learning from patients: Let's get real in medical education. *Advances in Health Sciences Education* 13:89-107
- Boshuizen, H. P. A., & Schmidt, H. G. (2008). The development of clinical reasoning expertise; Implications for teaching. In: J. Higgs, M. Jones, S. Loftus, & N. Christensen (Eds.), *Clinical reasoning in the health professions*, (3rd comp. rev. ed.). Oxford: Butterworth-Heinemann/Elsevier.
- Bowker, G., C. and Star, S. (1999) *L. Sorting Things Out: Classification and Its Consequences (Inside Technology)*. The MIT Press.
- Braidman I, Regan M, Cowie B (2008) Online reflective learning supported by student facilitators. *Medical Education* 42:528-529
- Brown, J. S. & Adler, R. P. (2008) Minds on fire; Open education, the long tail, and Learning 2.0. *Educause Review*, 16-20, 22, 24, 26, 28, 30, 32.
- Brown, J.S. & Duguid, P. (1991). Organizational Learning and Communities-of-Practice: Toward a Unified View of Working, Learning, and Innovation, *Organization Science*, 2 (1), 40-57.
- Burek, G., Pietsch, C., and De Roeck, A. (2007). SVO triple based Latent Semantic Analysis for recognising textual entailment. *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing (WTEP). 45th Annual Meeting of the Association of Computational Linguistics (ACL 2007)*, Prague. 113-118. 2007.
- Centre for Recording Achievement ePortfolios briefing paper. Retrieved November, 2008 from <http://www.recordingachievement.org/downloads/eportfolios-briefing-paperSGRW11.doc>.
- Charlin, B., Boshuizen, H. P. A., Custers, E., & Feltovich, P. J. (2007). Scripts and clinical reasoning. *Medical Education*, 41(12), 1178-1184(7).
- Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, 63, 1-51.
- Clariana, R., Koul, R., and Salehi, R. (2006). The criterion-related validity of a computer-based approach for scoring concept maps. *International Journal of Instructional Media*, 33(3), 317-325.

- Cover, T. and Hart P. (1967) Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21– 27.
- Dornan T., & David T. (2000) Adult Learning and Continuing Education *Diabetic Medicine* 17: 78-80
- Dysthe, O. (1996) 'The Multivoiced Classroom: Interactions of Writing and Classroom Discourse'. *Written Communication*, 13.3: 385-425
- Eraut, M. (1994). *Developing professional knowledge and competence*. London: RoutledgeFalmer.
- Ericsson, K. A. (ed., 1996). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games*. Mahwah, NJ: Erlbaum.
- Ericsson, K. A. (1999). Creative expertise as superior reproducible performance: innovative and flexible aspects of expert performance, *Psychological Inquiry* 10, 4, 329–361
- Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine*, 10, S70-S81.
- Ertmer, P. A., Newby, T. J. (1993). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly*, 6 (4), 50-70.
- Galley M., and McKeown K. (2003). Improving word sense disambiguation in lexical chaining. In *Proceedings of 18th International Joint Conference on Artificial Intelligence (IJCAI)*, Acapulco, Mexico.
- Gangemi, A., Navigli, R., and Velardi, P. (2003). The OntoWordNet Project: extension and axiomatization of conceptual relations in WordNet. R. Meersman et al. (Eds.), *Conference on Ontologies, Databases and Applications of SEmantics (ODBASE 2003)*, Catania, (Italy), Springer.
- Garrison DR, Anderson T, Archer W (2000) Critical inquiry in a text-based environment: computer conferencing in higher education. . *The Internet and Higher Education* 2:87 - 105.
- Goldsmith, T. E., Johnson, P. J., and Acton, W. H. (1991). Assessing structural knowledge. *Journal of Educational Psychology*, 83, 88–96.
- Guarino, N. (2000). Invited Mini-course on Ontological Analysis and Ontology Design. First Workshop on Ontologies and lexical Knowledge Bases - OntoLex 2000. Sozopol, Bulgaria.
- Guba EG and Lincoln YS (2005), “Paradigmatic controversies, contradictions and emerging confluences” in *The Sage Handbook of Qualitative Research*, ed. Denzin NK and Lincoln YS, Sage Publications Ltd
- Hatano, G. & Inagaki, K. (1986). Two courses of expertise. In: H. W. Stevenson, H. Azuma, & K. Hakuta (Eds). *Child development and education in Japan* (pp. 262-272). New York: W H Freeman.
- Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research*, 66, 99-136.
- Hobbs, J. (1985). On the coherence and structure of discourse. CSLI TR 85-37. Stanford.
- Ifenthaler, D., and Seel, N. M. (2005). The measurement of change: learning-dependent progression of mental models. *Technology, Instruction, Cognition and Learning*, 2(4), 317-336.
- Institute for Human and Machine Cognition. (2004). Cmap tools: University of West Florida.
- Jeong, A. (2008). jMap v. 104. Available at <http://dev2244801.sp01.fsu.edu/ExcelTools/jmap>
- Jonassen, D. H., Beissner, K., and Yacci, M. (1993a). *Structural knowledge: techniques for representing, conveying, and acquiring structural knowledge*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Jonassen, D. H., McAleese, T. M. R. & Duffy, T. M. (1993b). A Manifesto for a constructivist approach to technology in higher education. In Duffy, T. M., Lowyck, J. & Jonassen, D. H. (Eds.) *Designing environments for constructive learning: Implications for instructional design and the use of technology*, Heidelberg, FRG: Springer-Verlag
- Jonassen, D., Reeves, T., Hong, N., Harvey, D., and Peter, K. (1997). Concept mapping as cognitive learning and assessment tools. *Journal of interactive learning research*, 8(3/4), 289-308.
- Kalz, M., Van Bruggen, J., Rusman, E., Giesbers, B., and Koper, R. (2007). Positioning of learners in Learning Networks with content, metadata and ontologies. *Interactive Learning Environments*, 15(2), 191–200.

- Kelly GA (1955) *The Psychology of personal constructs*, vol. 1 and 2. Norton, New York.
- Kennedy, A. & Inkpen, D. (2006). Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence*. 22 (2), 110-125.
- Kim, S. & Hovy, E. (2006). Extracting opinions, opinion holders, and topics expressed in online news media text. In *Proceedings of ACL/COLING Workshop on Sentiment and Subjectivity in Text*, Sydney, Australia.
- Kim, D. K., Kim, M. and Park, H. (2008) Linearized suffix tree: an efficient index data structure with the capabilities of suffix trees and suffix arrays. *Algorithmica*, 52(3):350–377.
- Knowles, M. S. (1975) *Self-Directed Learning. A guide for learners and teachers*, Englewood Cliffs: Prentice Hall/Cambridge.
- Landauer, T. K. & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.
- Lave, J., and Wenger, E. C. (1991). *Situated Learning: legitimate peripheral participation*. New York: Cambridge University Press.
- Liu, B. (2008). *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. Springer.
- Lown N, Davies I, Cordingley L, Bundy C, Braidman I (2009) Development of a method to investigate medical students' perceptions of their personal and professional development. *Adv Health Sci Educ Theory Pract* 14:475-86.
- Merrifield, J., McIntyre, D. and Osaigbovo, R. (2000). Mapping APEL: Accreditation of prior experiential learning in English higher education. London: Learning from Experience Trust.
- Mieg, H. A. (2009). Two factors of expertise? Excellence and professionalism of environmental experts. *High Ability Studies*, 20(1), 91-115.
- Mihalcea R. (2007). Word Sense Disambiguation, *Encyclopedia of Machine Learning*, Springer.
- Moilanen, K. and Pulman, S. (2007). Sentiment Composition. In *Proceedings of Recent Advances in Natural Language Processing (RANLP 2007)*. Borovets, Bulgaria, 378-382.
- Nesbit, J., and Adesope, O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, 76, 413-448.
- Newman DR, Johnson C, Webb B, Cochrane C (1997) Evaluating the quality of learning in computer supported co-operative learning. *Journal of the American Society for Information Science* 48:484 – 496
- Nokes, T. J., & Ohlsson, S. (2003). Declarative transfer from a memory task to a problem solving task. *Cognitive Science Quarterly*, 3, 259-296.
- Nonaka, I., Nayana, R., & Konno, N. (2000). SECI, Ba and Leadership: a unified model of dynamic knowledge creation. *Long Range Planning*, 30, 5-34.
- Novak, J. D. (1998). *Learning, creating and using knowledge: concept maps as facilitative tools in schools and corporations*. Mahwah, NJ: Erlbaum.
- Norman, G. R., Eva, K., Brooks, L., & Hamstra, S. (2006). Expertise in medicine and surgery, In: K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance*, (pp 344-347). Cambridge: Cambridge University Press.
- O'Neill, P. A., Willis, S. C., and Jones, A. (2002). A model of how students link problem-based learning with clinical experience through "Elaboration". *Academic Medicine*, 77, 552–561.
- Oberholzer, R., Egloff, S., Ryf, S., and Läge, D. (2008) ProDax. Datenauswertung im Bereich der Skalierung. Benutzerhandbuch. Available at <http://www.prodax.ch/ProDaX-Handbuch.pdf>
- Pang, B and Lee, L. (2004). A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In *Proceedings of ACL 2004*, 271-278.
- Peters, S. (2005). Infomap NLP software: an open-source. Package for natural language processing. Available at <http://infomap-nlp.sourceforge.net/>
- Pirnay-Dummer, P. (2006). Expertise and model building. MITOCAR. Unpublished doctoral dissertation. University of Freiburg, Freiburg, Germany.

- Rimon, M. (2005). Sentiment Classification: Linguistic and Non-linguistic Issues. In Proceedings of Israel Association for Theoretical Linguistics, 21.
- Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist*, 24, 113–142.
- Salton, G. and Buckley, C. (1988) Term-weighting approaches in automatic text retrieval. In Information Processing and Management, pages 513–523.
- Sargeant, J Mann, K, Sinclair, D Ferrier S et al (2006) Learning in practice: Experiences and perceptions of high-scoring physicians *Academic Medicine* 81: 655 – 660.
- Schauer, H. (2000). Referential structure and coherence structure. TALN 2000. Lausanne, Switzerland.
- Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Academic Medicine*, 65, 611–621.
- Schön, D. (1987). *Educating the reflective practitioner; toward a new design for teaching and learning in the professions*. San Francisco: Jossey-Bass.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15 (2), 4–14.
- Shute, V. (2008). Focus on Formative Feedback. *Review of Educational Research*, 78 (1), 153–189.
- Shute, V. J., Jeong, A. C., Spector, J. M., Seel, N. M., and Johnson, T. E. (submitted). Model-Based Methods for Assessment, Learning, and Instruction: Innovative Educational Technology at Florida State University. In Yearbook Educational Media and Technology.
- Schvaneveldt, R. W. (1990). *Pathfinder associative networks: Studies in knowledge organization*. Nordwood, NJ: Ablex.
- Simov, K., Peev, Z., Kouylekov, M., Simov, E., Dimitrov M. and Kiryakov, A. (2001). CLaRK - an XML-based System for Corpora Development. In Proceedings of the Corpus Linguistics 2001 Conference, 558 -560
- Simov, K. and Osenova, P. (2007). Applying Ontology-Based Lexicons to the Semantic Annotation of Learning Objects. In Proc. Of RANLP 2007 workshop: Natural Language Processing and Knowledge Representation for eLearning Environments. Borovets, 26. September 2007.
- Simov, K. and Osenova, P. (2008). Language Resources and Tools for Ontology-Based Semantic Annotation. In Proc. at the Workshop ‘OntoLex 2008’ at LREC 2008.
- Smith, A. E., & Humphreys, M. S. (2006). Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping. *Behavior Research Methods*, 38(2), 262–279.
- Spector, J. M., and Koszalka, T. A. (2004). The DEEP methodology for assessing learning in complex domains (Final report to the National Science Foundation Evaluative Research and Evaluation Capacity Building). Syracuse, N.Y: Syracuse University.
- Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1992). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. In T. M. Duffy & D. H. Jonassen (Eds.), *Constructivism and the technology of instruction: A conversation*. Hillsdale, NJ: Lawrence Erlbaum Associates, 57–76.
- Stahl, G. (2006) *Group Cognition: Computer Support for Building Collaborative Knowledge*. Cambridge: MIT Press.
- Sternberg, R. J. & Frensch, P. A. (1992). On being an expert: A cost-benefit analysis. In R. R. Hoffman (Ed.), *The psychology of expertise: Cognitive research and empirical AI* (pp 191–203). New York: Springer.
- Thijssen, J. G. L., & van der Heijden, B. I. J. M. (2003). Evaporated talent? Problems with talent development during the career. *International Journal of Human Resources Development and Management*, 3(2), 154–170.
- van Bruggen, J., Sloep, P., van Rosmalen, P., Brouns, F., Vogten, H., Koper, R., and Tattersall, C. (2004). Latent Semantic Analysis as a tool for learner positioning in learning networks for lifelong learning. *British Journal of Educational Technology*, 35 (6), 729–738.
- Van der Heijden, B. I. J. M. (2000). The development and psychological evaluation of a multi-dimensional measurement instrument of professional expertise, *High Ability Studies*, 11(1), 9–39.

- van Merriënboer, J. J. G., Clark, R., & de Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational Technology, Research and Development*, 50, 39-64.
- van Rosmalen, P., Sloep, P., Kester, L., Brouns, F., De Croock, M., Pannekeet, K., et al. (2008). A learner support model based on peer tutor selection. *Journal of Computer Assisted Learning*, 24(1), 74-86.
- Vosniadou, S., & Ortony, A. (1989). Similarity and analogical reasoning: A synthesis. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 1-17). Cambridge: Cambridge University Press.
- Vygotsky, L.S. (1978). *Mind in Society*. Cambridge, MA: Harvard University Press.
- Weggeman, M. C. D. P. (2000). *Kennismanagement: de praktijk* [Knowledge management: the practice]. Schiedam: Scriptum.
- Wenger E (2001) Supporting communities of practice: a survey of community-oriented technologies. Self-published report available at www.ewenger.com/tech.
- Wenger, E. C. (1998). *Communities of Practice, learning, meaning and identity*. New York: Cambridge University Press.
- Wertsch, J. V. (1991) *Voices of the mind: A sociocultural approach to mediated action*. Cambridge, MA : Harvard University Press
- Wiebe, J. (2002). Instructions for Annotating Opinions in Newspaper Articles. Department of Computer Science Technical Report TR-02-101, University of Pittsburgh, Pittsburgh, PA.
- Wolf, F. & Gibson, E. (2006). *Coherence in Natural Language. Data structures and applications*. A Bradford Book. The MIT Press.
- Woods, N. N. (2007). Science is fundamental: the role of biomedical knowledge in clinical reasoning. *Medical Education*, 41, 1173-1177.
- Woods, N. N., Brooks, L. R., & Norman, G. R. (2005). The value of basic science in clinical diagnosis: Creating coherence among signs and symptoms. *Medical Education*, 39, 107-112
- Woods, N. N., Brooks, L. R., & Norman, G. R. (2007). It all makes sense: Biomedical knowledge, causal connections and memory in the novice diagnostician. *Advances in Health Sciences Education*, 12, 417-426.
- Yamamoto, M. and Church, K.W. (2001) Using suffix arrays to compute term frequency and document frequency for all substrings in a corpus. *Comput. Linguist.*, 27(1):1-30.

-
- [1] Optionally the tokenizer could be set to eliminate such punctuation marks. The bullet point makes a good example here, however, due to its burstiness.
- [2] This definition and the following definitions are similar to those found in Abouelhoda et al 2004.
- [3] We assume here that the text is padded with unique beginning of string and end of string sentinels so that indexing at 0 or 7 makes sense.
- [4] In the process of experiments with the actual data we will refine this scale of values.
- [5] Notice that in this first exploration “expert knowledge” has been defined as the course content. The “expert knowledge” can be also seen as the knowledge a learner, who is considered by the tutor as an expert learner, has in a particular context.